

UNIVERSITY OF MIAMI

AN ANALYTICAL FRAMEWORK FOR SOFT AND HARD DATA FUSION: A
DEMPSTER-SHAFER BELIEF THEORETIC APPROACH

By

Thanuka L. Wickramarathne

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

August 2012

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE AUG 2012		2. REPORT TYPE		3. DATES COVERED 00-00-2012 to 00-00-2012	
4. TITLE AND SUBTITLE An Analytical Framework For Soft And Hard Data Fusion: A Dempster-Shafer Belief Theoretic Approach				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Miami, Coral Gables, FL, 33124				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The recent experiences of asymmetric urban military operations have highlighted the pressing need for incorporation of soft data, such as informant statements, into the fusion process. Soft data are fundamentally different from hard data (generated by physics-based sensors), in the sense that the information they provide tends to be qualitative and subject to interpretation. These characteristics pose a major obstacle to using existing multi-sensor data fusion frameworks, which are quite well established for hard data. Given the critical and sensitive nature of intended applications soft/hard data fusion requires a framework that allows for convenient representation of various data uncertainties common in soft/hard data, and provides fusion techniques that are robust, mathematically justifiable, and yet effective. This would allow an analyst to make decisions with a better understanding of the associated uncertainties as well as the fusion mechanism itself.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 237	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

©2012
Thanuka L. Wickramaratne
All Rights Reserved

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

AN ANALYTICAL FRAMEWORK FOR SOFT AND HARD DATA FUSION: A
DEMPSTER-SHAFER BELIEF THEORETIC APPROACH

Thanuka L. Wickramaratne

Approved:

Kamal Premaratne, Ph.D.
Professor of Electrical and Computer Engineering

M. Brian Blake, Ph.D.
Dean of the Graduate School

Manohar N. Murthi, Ph.D.
Associate Professor of Electrical and Computer Engineering

James W. Modestino, Ph.D.
Professor of Electrical and Computer Engineering

Miroslav Kubat, Ph.D.
Associate Professor of Electrical and Computer Engineering

Marco Pravia, Ph.D.
BAE Systems
Burlington, Massachusetts, USA

WICKRAMARATHNE, THANUKA L.

(Ph.D., Electrical and Computer
Engineering)

An Analytical Framework for Soft and Hard Data Fusion:
A Dempster-Shafer Belief Theoretic Approach

(August 2012)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Kamal Premaratne.

No. of pages in text: (213)

The recent experiences of asymmetric urban military operations have highlighted the pressing need for incorporation of *soft data*, such as informant statements, into the fusion process. Soft data are fundamentally different from *hard data* (generated by physics-based sensors), in the sense that the information they provide tends to be qualitative and subject to interpretation. These characteristics pose a major obstacle to using existing multi-sensor data fusion frameworks, which are quite well established for hard data. Given the critical and sensitive nature of intended applications, soft/hard data fusion requires a framework that allows for convenient representation of various data uncertainties common in soft/hard data, and provides fusion techniques that are robust, mathematically justifiable, and yet effective. This would allow an analyst to make decisions with a better understanding of the associated uncertainties as well as the fusion mechanism itself.

We present here a detailed account of an analytical solution to the task of soft/hard data fusion. The developed analytical framework consists of several main components: **(i)** a Dempster-Shafer (DS) belief theory based fusion strategy, **(ii)** a complete characterization of the Fagin-Halpern DS theoretic (DST) conditional notion which forms the basis of the data fusion framework, **(iii)** an evidence updating strategy for the purpose of consensus generation, **(iv)** a credibility estimation technique for validation of evidence, and **(v)** techniques for reducing computational burden associated with the proposed fusion framework.

The proposed fusion strategy possesses several intuitively appealing features, and satisfies certain algebraic and fusion properties making it particularly useful in a

soft/hard fusion environment. This strategy is based on DS belief theory which allows for convenient representation of uncertainties that are typical of soft/hard domains.

The Fagin-Halpern (FH) notion is perhaps the most appropriate DST conditional notion for soft/hard data fusion scenarios. It also forms the basis for our fusion framework. We provide a complete characterization of the FH conditional notion. This constitutes a strong result, that sets the foundation for understanding the FH conditional notions and also establishes the theoretical grounds for development of algorithms for efficient computation of FH conditionals. We also address the converse problem of determining the evidence that may have generated a given change of belief. This converse result can be of significant practical value in certain applications.

A consensus control strategy developed based on our fusion technique allows consensus analysis to be carried out in a multitude of applications that call for extended flexibility in uncertainty modeling. We provide a complete theoretical development of the proposed consensus strategy with rigorous proofs. We make use of these consensus notions to establish a data validation technique to assess credibility of evidence in the absence of ground truth. Credibility estimates can be used in fusion equations and also be used to estimate reliability of sources for subsequent fusion operations.

Computational overhead is one of the major obstacles associated with data fusion operations, especially in DS theoretic methods. We propose a graphical procedure and its associated message passing scheme for efficient computation of the conditionals, along with the theoretical bounds for computational costs. In addition, we propose a method based on statistical sampling techniques to approximate DST data models. This allows for efficient computational representations as well as further reductions in computational costs associated with DS theoretic fusion operations.

We have used several example scenarios throughout the presentation to clarify and validate the proposed notions and techniques. We conclude the dissertation by providing several guidelines for future research and summary of the work that is being presented.

... to my parents!

Acknowledgements

This monograph reports the culmination of four years of research, which would not have been possible without the help and support of many loved ones.

First and foremost, I extend my sincere gratitude and appreciation to my PhD advisor and chairman of the thesis committee, Professor Kamal Premaratne, for his guidance, support and suggestions throughout this work. Not only did he teach me everything I know about research, but he has also given me great advice and support in my personal life over the years. My special thanks also go to Professors Manohar N. Murthi, Miroslav Kubat and James W. Modestino of the Department of Electrical and Computer Engineering, and Dr. Marco Pravia of BAE Systems, for accepting the appointment to the dissertation committee and for their helpful suggestions and support. The fruitful discussions I had with numerous professors from the Departments of Electrical and Computer Engineering, and Mathematics were invaluable in completing this work. I also like to thank the office staff of Department of Electrical and Computer Engineering for their kind support. Encouragement and support I received from my friends and colleagues is highly appreciated.

During various times of my graduate research work, I was fortunate to have been supported by research grants from multiple federal funding agencies. This support I received from the US National Science Foundation (NSF) Grant #0513702, US Department of Defense (DoD) Grant #W912HQ-09-C-0001, US Office of Naval Research (ONR) Grants #N00014-10-1-0140 and #N00014-11-1-0493, is gratefully acknowledged. In particular, I would like to thank Dr. Ivy Estabrooke, the Program Manager of the ONR, Human, Social, Culture, and Behavior (HSCB) Modeling Thrust, for her encouragement, guidance and input while the research described in this dissertation was conducted.

Finally, I extend my utmost gratitude to my parents, my brother Kanishka, my sister Gayani, and last but not least, my wife Thanushi for their encouragement, patience and love, which made this work a reality.

THANUKA L. WICKRAMARATHNE

University of Miami

August 2012

Table of Contents

	Page
LIST OF FIGURES	x
LIST OF TABLES	xii
LIST OF SYMBOLS	xiv
1 INTRODUCTION	1
1.1 Overview	1
1.2 Motivation	2
1.3 Challenges	3
1.3.1 Accommodating Data Imperfections	3
1.3.2 Accommodating Source Differences	4
1.3.3 Adaptability of Fusion Operations	5
1.3.4 “Well-founded” Fusion Framework	6
1.3.5 Computational Burden	7
1.4 Contributions	7
1.4.1 Conditional Core Theorem	8
1.4.2 Conditional Approach to Evidence Fusion	9
1.4.3 Consensus in DST Fusion Environments	9
1.4.4 Credibility Estimation in the Absence of Ground Truth	10
1.4.5 Efficient Computation of Conditionals	10
1.4.6 Approximations Via Statistical Sampling	11
1.5 Organization of the Dissertation	11
I PRELIMINARY NOTIONS	14
2 MULTI-SENSOR DATA FUSION	15

2.1	Data Fusion	15
2.1.1	Processing Architectures	16
2.2	Fusion Models	17
2.3	Challenges for Multi-sensor Fusion	18
3	UNCERTAIN DATA PROCESSING	21
3.1	A Taxonomy of Uncertain Data	22
3.2	Approaches to Handling Uncertainty	24
3.2.1	Bayesian Probability Theory	25
3.2.2	Possibility Theory	27
3.2.3	Dempster-Shafer Theory	28
3.3	A Comparison of Popular Formalisms	29
3.4	Detailed Review of Dempster-Shafer Theory	30
3.4.1	Basic Notions	31
3.4.2	Conditional Notions	32
3.4.3	Evidence Combination	33
3.4.4	Evidence Updating	36
3.4.5	Evidence Fusion: Non-identical Scopes Case	37
3.4.6	Evidence Discounting	39
3.4.7	Distance Measures	40
3.5	Chapter Summary	40
II	SOFT/HARD DATA FUSION FRAMEWORK	41
4	DST CONDITIONAL NOTIONS	42
4.1	Which Conditional Notion? And Why ?	43
4.2	Characterization of the Conditional Core	45
4.2.1	Preliminaries	45
4.2.2	Conditional Core Theorem	47
4.2.3	Implications of the Conditional Core Theorem	50
4.2.4	Converse of the Conditional Core Theorem	51
4.3	Application Example	52
4.4	Chapter Summary	56
5	CONDITIONAL APPROACH TO DATA FUSION	57
5.1	Ported Conditional Notions	58
5.1.1	Ported Conditional Belief Function	59
5.1.2	Ported Conditional Plausibility Function	60
5.1.3	Ported Conditional Mass Function	61
5.2	Conditional Fusion Equation	62
5.2.1	Properties of Conditional Fusion Equation	65
5.2.2	Evidence Updating	69

5.2.3	Fusion of Multiple Bodies of Evidence	70
5.3	Selection of Parameters	73
5.3.1	Fusion Scope	74
5.3.2	Conditioning Sets	74
5.3.3	Linear Combination Weights	74
5.4	An Illustrative Example	79
5.5	Application Examples	84
5.5.1	Natural Language Processing	84
5.5.2	Underwater Objects Characterization	89
5.6	Chapter Summary	95
6	CONSENSUS IN A FUSION ENVIRONMENT	96
6.1	Overview	97
6.1.1	Agent Interactions	97
6.1.2	Consensus	100
6.1.3	Consensus Protocols	101
6.2	Paracontractions View of Consensus Protocols	104
6.2.1	Paracontracting Operators	104
6.2.2	Non-Linear Asynchronous Iterations	105
6.2.3	Confluence and Regulatory Assumptions	107
6.2.4	Convergence Criteria	110
6.2.5	Non-Linear Asynchronous Consensus Protocols	111
6.3	Consensus in DST Fusion Environments	115
6.3.1	Consensus in a DST Fusion Environment	116
6.3.2	Asynchronous Consensus Protocols for DST BoEs	118
6.3.3	Existence of Consensus: Proof of Convergence	121
6.4	Chapter Summary	125
7	DATA VALIDATION FOR ROBUST FUSION	127
7.1	Overview	128
7.2	Credibility Estimation	128
7.2.1	Conflict-based Credibility Estimation	129
7.2.2	Consensus-based Credibility Estimation	129
7.3	An Illustrative Example	130
7.4	Chapter Summary	133
III	COMPUTATIONAL OPTIMIZATIONS	134
8	EFFICIENT COMPUTATION OF CONDITIONALS	135
8.1	Overview	136
8.2	Direct Computation of Conditional Masses	137
8.2.1	Polytree Representation of Conditional Core	138

8.2.2	Conditional Masses via Message-Passing	139
8.3	Computational Gains	140
8.4	Experiments	143
8.4.1	Data Generation	143
8.4.2	Results and Analysis	145
8.5	Chapter Summary	149
9	DST CORE APPROXIMATIONS	151
9.1	Overview	152
9.2	Approximations of the Core	153
9.2.1	The Bayesian Approximation (BA)	153
9.2.2	The k - ℓ - x Method ($k\ell x$)	154
9.2.3	Summarization Method (SM)	154
9.2.4	D1 Approximation (D1)	155
9.3	Monte Carlo Core Approximation (MCCA)	156
9.3.1	Problem Formulation	156
9.3.2	Algorithm	157
9.3.3	Selection of an Appropriate Objective Function	160
9.4	Experiments	162
9.5	Chapter Summary	168
IV	EPILOGUE	173
10	FUTURE RESEARCH DIRECTIONS	174
10.1	Distributed Fusion	174
10.2	Implication Rules as Conditionals	176
10.3	Reliability/Credibility Estimation	177
11	CONCLUSION	179
A	PROOFS	183
A.1	Chapter 4 Proofs	183
A.2	Chapter 5 Proofs	196
A.3	Chapter 6 Proofs	198
	Appendices	183
	BIBLIOGRAPHY	204

List of Figures

1.1	Representative information elements in an urban operation	2
2.1	JDL Fusion Model (1992 Revision) [1]	18
2.2	A Complex Fusion Scenario [1]	19
3.1	Conditional Approach to Evidence Updating	39
4.1	A set theoretic interpretation of $\text{in}(A)$ and $\text{out}(A)$ in Definition 11. . .	46
5.1	Fusion Results of Example 5.4	82
5.2	Updating scheme used in the example	87
5.3	Variation of Shadow Geometry	91
5.4	Preprocessed Optical Images	92
6.1	Discrete-event based time	98
6.2	Spatial connectivity among agents in multi-agent system	100
6.3	A dynamic partially-connected multi-agent network	101
6.4	Interaction and iteration graphs	109
6.5	Example 15: An illustration of graph composition	114
6.6	A Rational Consensus in DST Sense	117

7.1	Convergence of Evidence BoEs	131
8.1	Polytree corresponding to $\mathfrak{F}_{\Theta A}$ in Example 1.	138
8.2	Variation of $ \mathfrak{F}_{\Theta A} $ versus $ A / \Theta $ and ρ , where $ \Theta = 12$	144
8.3	Variation of $ \mathfrak{F}_{\Theta A} $ versus $ \Theta $ for different $ A / \Theta $, with $\rho = 1/16$	145
8.4	$\mathcal{T}_{std}/\mathcal{T}_{thm}$ versus $ A /\Theta$ for different ρ , with $ \Theta = 12$	146
8.5	$\mathcal{T}_{std}/\mathcal{T}_{thm}$ versus ρ for different $ A $, with $ \Theta = 12$	147
8.6	$\mathcal{T}_{std}/\mathcal{T}_{thm}$ versus smaller ρ for different $ A $, with $ \Theta = 12$	148
8.7	$\mathcal{T}_{std} - \mathcal{T}_{thm}$ versus ρ for different $ A $, with $ \Theta = 12$	149
8.8	\mathcal{T}_{std} and \mathcal{T}_{thm} versus $ \Theta $, for different $ A $, with $ \mathfrak{F}_{\Theta} = 4$	150
9.1	MCCA-generated approximation: <i>Error1</i> Vs. $ \mathfrak{F}'_{\Theta} $	165
9.2	MCCA-generated approximation: <i>Error1</i> Vs. N_s	166
9.3	MCCA-generated approximation: <i>Error2</i> and <i>Error3</i> Vs. $ \mathfrak{F}'_{\Theta} $	167
9.4	MCCA-generated approximation: <i>RMS</i> and <i>MAE</i> Vs. $ \mathfrak{F}'_{\Theta} $	168
9.5	<i>Error1</i> of $\mathcal{E}_{\Theta_X}[i]$, $i = 1, \dots, 5$, the i -th DCR-fused BoE.	170
9.6	<i>Error2</i> and <i>Error3</i> of $\mathcal{E}_{\Theta_X}[i]$, $i=1, \dots, 5$, the i -th DCR-fused BoE.	171
9.7	<i>RMS</i> and <i>MAE</i> of $\mathcal{E}_{\Theta_X}[i]$, $i = 1, \dots, 5$, the i -th DCR-fused BoE.	172

List of Tables

3.1	A Classification of Uncertain Data Types	23
3.2	Approaches to Handling Uncertainty in Data	25
3.3	Uncertainty Output types for different formalisms	30
4.1	Explicit conditional mass computation for Example 1	49
4.2	Available Sensor Reports	53
4.3	Fused BoE Generated From the Sensor Reports	54
4.4	Conditional Core $\mathfrak{F}_{\Theta A}$ Corresponding to $A = nsh \times B_2B_4$	55
5.1	Computation Example of Ported Conditional Beliefs	60
5.2	Ported Conditional Belief Values for Example 4.	63
5.3	Fusion of Contradictory Evidence: Example from [2]	68
5.4	Fusion of Contradictory Evidence: Example from [3]	68
5.5	Summary of Fusion Properties of Commonly Used Combination Rules	69
5.6	Example 5.4: Evidence Gathered from Soft and Hard Sensor Suites .	80
5.7	Fused Results Corresponding to Example 5.4	81
5.8	Evidence gathered from WS_1 , WS_2 and DB_3	85
5.9	Evidence models of WS_1 , WS_2 and DB_3	86
5.10	Final fusion results.	88

5.11	Target Characteristics	90
5.12	Mass Assignment Models for Optical and Sonar Sensors	93
5.13	Mass Models of Fused Evidence BoEs	94
7.1	Estimated credibility measures of the BoEs.	132
8.1	Some Implementation Specific Constants	141
9.1	Core Approximation of Example 17	159
9.2	Conditional Core Corresponding to $A = (abcdefghij)$	160
9.3	Core Approximation of Example 18	161
9.4	Cardinality of the Core of the DCR-Fused BoE $\mathcal{E}_\Theta[k]$	169
10.1	Numerical assessments of truthfulness of evidence	177
10.2	Context classification of incidents	178

List of Symbols

Abbreviations: ---

klx	The k - ℓ - x Approximation method	§ 9.2.2	p. 154
TL	Threat Level		p. 82
BA	Bayesian Approximation method	§ 9.2.1	p. 153
BBA	Basic Belief Assignment		p. 31
BoE	Body of Evidence		p. 31
cCFE	cautious CFE	Def. 21	p. 76
CCG	Conditional Core Generator		p. 51
CCT	Conditional Core Theorem		p. 47
cCUE	cautious CUE		p. 78
CFE	Conditional Fusion Equation		p. 62
COMINT	COMmunications INTelligence		p. 2
CUE	Conditional Update Equation		p. 69
D1	D1 Approximation method	§ 9.2.4	p. 155
DCR	Dempster's Combination Rule		p. 34
DP1	Dubois and Prade's Rule #1		p. 35
DP2	Dubois and Prade's Rule #2		p. 35

DS	Dempster Shafer		p. 7
DST	Dempster Shafer Theoretic		p. 8
FH	Fagin-Halpern		p. 33
FoD	Frame of Discernment		p. 28
GEOINT	GEOspatial INTelligence		p. 1
GT	ground truth		p. 115
HUMINT	HUMan INTelligence		p. 1
KB	Knowledge Base		p. 4
LHS	Left Hand Side		p. 64
MAP	Maximum a-posteriori Probability		p. 26
MASINT	Measurements And Signal INTelligence		p. 1
MCCA	Monte Carlo Core Approximation	§ 9.3	p. ??
ML	Maximum Likelihood		p. 26
MURI	Multi University Research Initiatives		p. 2
NLP	Natural Language Processing		p. 84
OSMINT	Open Source INTelligence		p. 2
PCM	Plausibilities Correction Method		p. 37
PCR	Partial Conflict Redistribution Rule		p. 36
rCFE	receptive CFE	Def. 20	p. 75
RCR	Robust Combination Rule		p. 36
rCUE	receptive CUE		p. 78
RHS	Right Hand Side		p. 64
SIGINT	SIGnal INTelligence		p. 1
SM	SM Approximation method	§ 9.2.3	p. 154
YGR	Yager's Rule		p. 35

Preliminary Notations:

Ω	sample space	§ 3.2.1	p. 25
Φ	σ -algebra on Ω	§ 3.2.1	p. 25
$P(\cdot)$	probability mass function (pmf)	§ 3.2.1	p. 25
$P(\cdot E)$	conditional pmf given evidence E	Eqn. (3.1)	p. 26
Π_{pos}	possibility measure	§ 3.2.2	p. 27
\mathcal{N}_{pos}	necessity measure	§ 3.2.2	p. 27
Θ	frame of discernment (FoD)	§ 3.4.1	p. 31
\mathfrak{F}_{Θ}	core	§ 3.4.1	p. 31
$m_{\Theta}(\cdot)$	mass function	Def. 1	p. 31
\mathcal{E}_{Θ}	body of evidence (BoE)	§ 3.4.1	p. 31
$\mathcal{E}_{\Theta}[k]$	time/event-indexed BoE at index k	§ 3.4.4	p. 36
$\text{Bl}_{\Theta}(\cdot)$	belief function	Def. 2	p. 32
$\text{Pl}_{\Theta}(\cdot)$	plausibility function	Def. 2	p. 32
$\hat{\mathfrak{F}}_{\Theta}$	set of propositions B s.t. $\text{Bl}(B) > 0$	§ 3.4.1	p. 32
$\text{BetP}(\cdot)$	pignistic probability	Eqn. (3.8)	p. 32
$\text{Bl}_{\Theta}(\cdot A)$	conditional belief function given A	Thm. 2	p. 33
$\text{Pl}_{\Theta}(\cdot A)$	conditional plausibility function given A	Thm. 2	p. 33
$m_{\Theta}(\cdot A)$	conditional mass function given A	Eqn. (3.12)	p. 33
$\mathfrak{F}_{\Theta A}$	conditional core w.r.t. event A	§ 3.4.2	p. 33
$\mathcal{E}_{\Theta_1} \bowtie \mathcal{E}_{\Theta_2}$	fusion of \mathcal{E}_{Θ_1} and \mathcal{E}_{Θ_2}	§ 3.4.3	p. 33
$\mathcal{E}_{\Theta_1} \triangleleft \mathcal{E}_{\Theta_2}$	update of \mathcal{E}_{Θ_1} with \mathcal{E}_{Θ_2}	§ 3.4.4	p. 36
$\mathcal{E}_{\Theta_1} \oplus \mathcal{E}_{\Theta_2}$	DCR-based combination of \mathcal{E}_{Θ_1} and \mathcal{E}_{Θ_2}	Def. 4	p. 34
$m_{\Theta_1 \cap \Theta_2}(\cdot)$	conjunctive rule on $m_{\Theta_1}(\cdot)$ and $m_{\Theta_2}(\cdot)$	Def. 3	p. 34
$m_{\Theta_1 \cup \Theta_2}(\cdot)$	disjunctive rule on $m_{\Theta_1}(\cdot)$ and $m_{\Theta_2}(\cdot)$	Def. 3	p. 34
$\text{dist}(\cdot, \cdot)$	JGB distance measure on DST BoEs	Def. 10	p. 40

$\text{in}(A)$	<i>inner set</i> of conditioning set A	Def. 11	p. 45
$\text{out}(A)$	<i>outer set</i> of conditioning set A	Def. 11	p. 45
$\text{IN}(A)$	arbitrary unions of elements in $\text{in}(A)$	Def. 11	p. 45
$\text{OUT}(A)$	arbitrary unions of elements in $\text{out}(A)$	Def. 11	p. 45
$\mathcal{S}(A; B)$	cumulative mass	Def. 12	p. 46
$K_{\Theta_2}(\Theta_1 A)$	normalizing constant of ported operators	§ 5.1	p. 58
$\text{Bl}_{\Theta_1}^{(\Theta_2)}(\cdot A)$	ported conditional belief function	Def. 13	p. 59
$\text{Pl}_{\Theta_1}^{(\Theta_2)}(\cdot A)$	ported conditional plausibility function	Def. 14	p. 60
$m_{\Theta_1}^{(\Theta_2)}(\cdot A)$	ported conditional mass function	Def. 15	p. 61
$\beta_i(\cdot)$	non-linear combination weights (CFE))	Def. 16	p. 62
$\alpha_i[k]$	CUE parameter of \mathcal{E}_{Θ_i} for update $\mathcal{E}_{\Theta_i} \triangleleft \mathcal{E}_{\Theta_j}$	Def. 17	p. 69
$\mathcal{E}_{\{1:i\}}$	iteratively fused BoE	§ 5.2.3.2	p. 71
\mathcal{K}_i	normalized relative importance measure	§ 5.3.3	p. 74
\mathcal{D}_i	unions of conditioning sets	Eqn. (5.22)	p. 75
\mathcal{S}_X	sensor suite X		p. 79
$Sh_{\theta_i}(\alpha_{in})$	shadow of target θ_i at angle α_{in}		p. 90
$Sh_{O_\Theta}(\alpha_{in})$	shadow of object O_Θ at angle α_{in}		p. 91
t_k	discrete event-based time	§ 6.1.1	p. 97
k	discrete event-based time index	§ 6.1.1	p. 97
\mathcal{A}_i	i^{th} interacting agent	§ 6.1.1	p. 97
\mathcal{N}	set of all interacting agents	§ 6.1.1	p. 97
$Q_{i,j}$	j^{th} interaction topology used by agent \mathcal{A}_i	§ 6.1.1	p. 99
\mathcal{Q}	set of all interaction topologies	§ 6.1.1	p. 99
$e_{i,j}$	information exchange from \mathcal{A}_i to \mathcal{A}_j	§ 6.1.1	p. 99
$E[k]$	set of edges modeling agent interactions	§ 6.1.1	p. 99

$G[k]$	agent interaction model at k	§ 6.1.1	p. 99
$\mathbf{x}_i[k]$	state of agent \mathcal{A}_i at k	Def. 22	p. 100
\mathbb{D}	domain of agent states	§ 22	p. 100
F^{ij}	state update function corresponding to $Q_{i,j}$	Def. 23	p. 102
$s^\ell[k]$	iteration lag	Def. 23	p. 102
\mathbb{N}_0	set of non-negative integers	Def. 23	p. 102
ξ	fixed point	Def. 24	p. 104
$fix(F)$	set of fixed points of operator F	Def. 24	p. 104
ζ	common fixed point	Def. 24	p. 104
\mathbb{N}	natural numbers	Def. 26	p. 105
\mathcal{F}	pool of operators	Def. 26	p. 105
\mathbb{I}	set of indices corresponding to \mathcal{F}	Def. 26	p. 105
\mathbf{X}	vector of elements of \mathbb{D}	Def. 26	p. 105
\mathcal{X}_0	set of initial conditions	Def. 27	p. 105
\mathcal{S}	sequences of m_i tuples	Def. 27	p. 105
\mathcal{V}	set of vertices of digraph $(\mathcal{V}, E_{\mathcal{S}})$	Def. 29	p. 109
$E_{\mathcal{S}}$	set of edges of digraph $(\mathcal{V}, E_{\mathcal{S}})$	Def. 29	p. 109
\mathcal{G}	set of all directed graphs with vertex set \mathcal{N}	Def. 31	p. 113
\mathcal{E}_{Θ}^t	Ground Truth BoE	Def. 33	p. 116
$\hat{\mathcal{E}}_{\Theta}^t$	Estimate of Ground Truth	Def. 33	p. 116
\mathcal{E}_{Θ}^*	Consensus BoE	Def. 33	p. 116
\mathfrak{E}_{Θ}	set of all possible BoEs on Θ	Def. 34	p. 118
F_{\triangleleft}^i	CUE based operator on DST BoEs	Def. 34	p. 118
$Cr_{cf}(\mathcal{E}_{\Theta_i})$	conflict-based credibility of \mathcal{E}_{Θ_i}	Def. 36	p. 129
$Cr_{cf1}(\mathcal{E}_{\Theta_i})$	conflict-based credibility variant-1	Def. 36	p. 129
$Cr_{cf2}(\mathcal{E}_{\Theta_i})$	conflict-based credibility variant-2	Def. 36	p. 129
$Cr_{con}(\mathcal{E}_{\Theta_i})$	consensus-based credibility of \mathcal{E}_{Θ_i}	Def. 37	p. 129

X_i	node corresponding to focal element B_{X_i}	§ 8.2.1	p. 138
\mathcal{X}	set of nodes corresponding to $\mathfrak{F}_{\Theta A}$	§ 8.2.1	p. 138
\mathbf{Ch}_{X_i}	set of child nodes of node $X_i \in \mathcal{X}$	§ 8.2.1	p. 138
\mathbf{Pa}_{X_i}	set of parent nodes of node $X_i \in \mathcal{X}$	§ 8.2.1	p. 138
\mathbf{Des}_{X_i}	set of descendent nodes of node $X_i \in \mathcal{X}$	§ 8.2.1	p. 138
$\mathbf{Des}_{X_i}^*$	represents $\mathbf{Des}_{X_i} \cup \{X_i\}$	§ 8.2.1	p. 138
\mathbf{S}_{X_i}	collection of focal elements as defined in	Eqn. 8.4	p. 138
\mathbf{F}_{X_i}	collection of focal elements as defined in	Eqn. 8.4	p. 138
$\mathcal{S}_{X_i}^{(F)}$	sum of masses of focal elements	Eqn. 8.5	p. 139
$\mathcal{S}_{X_i}^{(S)}$	sum of masses of focal elements	Eqn. 8.5	p. 139
$\mathcal{S}_{X_i}^{(A)}$	conditional mass of B_{X_i} given A	Eqn. 8.5	p. 139
$\mu_{X_j \rightarrow X_i}(X_i)$	message from node X_j	§ 8.2.2	p. 139
$\mathcal{T}_{id.ioa}$	cpu time to identify $\text{in}(A)$ and $\text{out}(A)$	§ 8.3	p. 141
$\mathcal{T}_{tr.bld}$	cpu time to build the polytree	§ 8.3	p. 141
$\mathcal{T}_{ps.msg}$	average cpu time per message		p. 141
$\mathcal{T}_{bp.ohd}$	average overhead without the CCT	§ 8.3	p. 141
$\mathcal{T}_{cp.std}$	time to compute $m_{\Theta}(\cdot A)$ without CCT	§ 8.3	p. 141
$\mathcal{T}_{cp.cct}$	time to compute $m_{\Theta}(\cdot A)$ with the CCT	§ 8.3	p. 141
$\mathcal{T}_{tr.pro}$	sum of $\mathcal{T}_{tr.bld}$ and $\mathcal{T}_{ps.msg}$	§ 8.3	p. 141
\mathcal{T}_{thm}	cpu time with CCT		p. 141
\mathcal{T}_{std}	cpu time with standard method		p. 141

Introduction

1.1 Overview

THE military, with its desire to fortify its technological stronghold, has always pushed the boundaries of science and engineering by driving researchers to tackle very challenging problems. The realm of *data fusion* —a discipline where refined position/identity estimates are sought by using techniques for association, correlation and combination of data [4]—is of no difference.

The recent experiences of asymmetric urban military operations [5] have highlighted the pressing need for incorporation of *soft data* (in the form of expert opinions, informant statements, tips etc.) into the fusion process [6]. *Multi-sensor data fusion*, i.e., fusion of data from multiple sensors, is quite well established for *hard data* [7–13]—such as GEOINT (GEOspatial INTelligence), SIGINT (SIGnal INTelligence) and MASINT (Measurements And Signatures INTelligence)—which are often generated by calibrated physical sensors with well-defined characteristics. In asymmetric warfare scenarios, soft sources (i.e., human-based sources) often provide the most crucial and perhaps the best complementary evidence to those provided by hard sensors [14, 15]. See Figure 2.1 for a classification of information types typical of such applications. However, human-generated soft data, such as HUMINT (HUMAN

INTelligence), OSINT (Open Source INTelligence) and COMINT (COMmunications INTelligence), are fundamentally different in the sense that the information they provide tend to be more qualitative and subjective to interpretation. Thus, a majority of the existing data fusion techniques, which are not capable of properly dealing with the uncertainties typical of soft data, become obsolete and are ineffective for soft/hard fusion applications [6, 16–18].

HUMINT	OSINT	SIGINT	GEOINT	MASINT
tips informant reports patrol debriefs links and relationships	political climate population sentiment culture TV/radio broadcasts websites	intercepted audio, imagery or video		
		<div> <div>SOFT</div> <div>HARD</div> </div>		
		signal frequency signal location	video and imagery spatial extent vehicle and building locations	seismic, magnetic, chemical, and other physical signatures identification event occurrence radar detections
coordinates	coordinates			

Figure 1.1: Representative information elements in an urban operation. Information elements are placed in the table according to generating source and classification between hard and soft information. [6]

1.2 Motivation

Defense-related research agencies have initiated soft/hard fusion research with a main focus on asymmetric urban warfare applications. The importance of this task is obvious from the increasing number of military-funded projects, challenges, competitions, MURIs (Multi University Research Initiatives), etc., dedicated to soft/hard fusion related work. Even though this soft/hard fusion related work as initiated in the defense establishment, its application in civil (i.e., non-military) scenarios is not far away. The key role that soft data can play in civil applications is already becom-

ing clear to the research community. The concept of *participatory sensing* [19], in which a community (of observers) is tasked to provide information for applications, such as urban planning and public health, is a clear example. In fact, a whole new paradigm of communication methods can be developed to link the military, local authorities and other agencies to establish a dynamic observational platform that makes use of humans, who are continuously adding content to the Internet, as sensors. To quote [20],

“... those billion amateurs are taking pictures of everything on the planet and placing the images on Flickr and other sites. There are thousands upon thousands of pictures of every known place, taken from all angles and under all lighting conditions. Researchers are now using those pictures to create three-dimensional images and panoramic vistas.”

Even though the fusion community has recently begun to look into soft/hard data issues [6, 14, 16, 17, 21–25], there is still a dearth of fusion frameworks catered specifically towards such data. Given the enormous potential for soft/hard data fusion techniques in both military and civil applications, a framework—which consists of robust, effective and yet mathematically justifiable fusion strategies that are capable of adequately representing and accommodating the types of imperfections inherent in soft/hard data—is very much in demand.

1.3 Challenges

Humans, in a sensing environment, do not behave similar to physical sensors. This fundamental difference in sources makes soft/hard data fusion a very challenging task.

1.3.1 C₁: Accommodating Data Imperfections

The types of uncertainties that can be handled by a fusion framework is often governed by the employed uncertainty handling formalism(s). A typical fusion en-

gine employs one or more uncertainty processing formalisms¹ for representation and processing of uncertain data.

For instance, if Bayesian probability is used, a *disjunction*, an uncertainty type typical of soft data, cannot be represented without making simplifying assumptions. To illustrate, consider a battlefield scenario, where a suite of soft and hard data sensors are employed for detection and identification of hostile units. A statement of the form: $\mathcal{SS}_1 \equiv \text{“I’m 90\% sure, it’s either a TANK or a TRUCK.”}$ is very typical of an evidence a witness may provide, for instance, about the `UNIT_TYPE` of a detected hostile unit. However, Bayesian inferencing cannot proceed² with $P(\text{UNIT_TYPE} = \text{TANK} \cup \text{TRUCK}) = 0.90$; but, on the other hand, a simplified model, such as $P(\text{UNIT_TYPE} = \text{TANK}) = P(\text{UNIT_TYPE} = \text{TRUCK}) = 0.45$, would not be able to capture what \mathcal{SS}_1 conveys adequately well. An uncertainty processing formalism suitable for soft/hard data fusion must be able to model these inherent imperfections with ease and still be able to capture the subtleties without making overly simplifying assumptions.

1.3.2 C₂: Accommodating Source Differences

Soft and hard sources often span different *scopes of expertise* (or simply *scopes*). For instance, in the battlefield scenario above, a hard sensor may be able to identify `fighter-jet(s)` with certain signatures, whereas a trained human military attachè may well be able to identify all different types of aircrafts (including `fighter-jet(s)`). Moreover, in a distributed sensor network setting, a node may be content in simply updating its existing *Knowledge Base (KB)* using the evidence that the node is routing to a neighboring node (i.e., “eavesdropping”). Thus, in a soft/hard data fusion scenario, the ability to fuse data from sources having different scopes, with only a moderate computational overhead, is a significant advantage. In addition, the ability to combine/update information without having to change one’s own scope could be

¹Chapter 3 provides a detailed discussion on uncertainty handling formalisms.

² $P(\text{UNIT_TYPE} = \text{TANK} \cup \text{TRUCK})$ must be distributed to $P(\text{UNIT_TYPE} = \text{TANK})$ and $P(\text{UNIT_TYPE} = \text{TRUCK})$ to proceed with inferencing, which requires simplifying assumptions to be made.

a particularly useful property to have in scenarios where the node possesses a scope with irrelevant content.

In addition to having different scopes of expertise, sources often possess different levels of reliability—a measure of being “*consistently good in performance*”/“*able to be trusted*” [26]; some sources may be more “*valuable*” (e.g., a warfare expert versus an informer) or may provide more *credible* information than others under given circumstances. Furthermore, one may not want to make significant changes to a KB built over time using evidence from multiple sources in favor of a single piece of evidence. The fusion framework must be capable of taking these factors into account. For instance, for updating the KB in the above battlefield scenario, one may want to assign more weight to evidence coming from a military attaché than a civilian; moreover, even if an incoming piece of evidence conflicts with the KB, if the information is given by a highly reliable source (e.g., a situation expert), one may still consider updating the existing KB.

1.3.3 C₃: Adaptability of Fusion Operations

One often encounters contradictory evidence in soft/hard fusion scenarios. If all sources are equally reliable, one may desire the fusion operation to be an aggregation of all available evidence. On the other hand, if the evidence from a source contradicts the current environment (e.g., the prevailing threat level), expert opinions and/or other information, the processing node may want to discard or lower the impact of this information.

In case of updating an existing KB, a processing node may need to be “cautious,” especially if the integrity of the KB is high and/or the incoming evidence partially or fully conflicts with the KB. In another instance, a source may be willing to be very “receptive” or even be prepared to completely ignore an existing KB (e.g., with a sudden change of the prevailing threat level). Thus, the fusion framework must provide this flexibility to adapt to various situations.

Moreover, A processing node may want to focus on evidence about a particular object. For instance, in the above battlefield scenario, if the base commander has confirmed evidence of a possible “air-attack,” s/he may want to refine incoming information to focus on evidence related to air-attacks and “filter out” the rest. So the fusion technique should be able to “condition” the incoming evidence to what is required under the given circumstances.

1.3.4 C₄: “Well-founded” Fusion Framework

Fusion engines are often used to obtain a better view or estimate of a scenario by fusing various inputs providing soft/hard evidence. An analyst, with a better understanding of the scenario, then takes final decisions or recommends necessary actions. This requires, especially in critical/sensitive application domains, the fusion frameworks to be “well-founded,” in the sense that the behavior of the fusion operations are fully characterized under all operating conditions; in other words, fusion engines cannot be “black-boxes.” Transparency of a fusion framework can be established by characterizing the fusion operations as a set of algebraic and fusion properties. For a given fusion technique, the validity and/or the usefulness of fusion operations are determined by the extent to which the so-called fusion properties are mathematically justifiable and applicable to the task at hand. Another aspect of transparency is that fused results must reflect or indicate the associated uncertainties that were fed into the fusion engine as part of the uncertainties associated with inputs; this allows, for instance, an analyst to make decisions with a full understanding of the associated uncertainties. Thus, in order to achieve transparency, data uncertainties must be propagated throughout the fusion process; and also, the fusion framework must be fully characterized via a set of properties applicable to various fusion tasks.

1.3.5 C₅: Computational Burden

Computational performance and the capacity for uncertainty processing (for example, within the modeling and inferencing processes) are often, in a very loose sense, “inversely” related. In other words, techniques that are computationally faster often do not offer the same level of uncertainty processing capacity of those techniques that are not as fast. For instance, certainty factor based methods do not offer the same modeling flexibility as that of Bayesian probabilistic methods, but they offer much faster inferencing capabilities. On the other hand, Dempster Shafer (DS) belief theoretic methods, which offer greater flexibility in uncertainty handling, even the task of conditioning can become computationally demanding in some situations [27].

However, for soft/hard data fusion, given the fact that majority of the targeted applications are in critical and sensitive domains (e.g., military and healthcare), it is imperative that one employs formalisms that are capable of capturing adequately well data uncertainties and propagating these uncertainties throughout the decision-making process so that the final analysis (or decision-making) can be done with a better awareness of the associated uncertainties. However, this comes at additional cost in terms of the computational overhead. This calls optimization and other techniques to make these uncertainty frameworks useful in more realistic applications.

1.4 Contributions

Three major formalisms—Bayesian probability theory, DS belief theory [28], and *possibility theory* based on *fuzzy reasoning* [29,30], are predominantly used for uncertain data handling³ in data fusion frameworks. DS belief theory, which provides a convenient framework [31] for working with a wide variety of data uncertainties [28,32,33], has emerged as one of the most dominant frameworks in a wide spectrum of application domains. These include, but are not limited to, signal processing [34–37], iden-

³Chapter 3 provides a discussion on uncertain data types and approaches for handling such data.

tification [38–40], remote sensing [41], and machine learning [42–48]. The increasing realization of the critical role soft evidence can play in the fusion and decision-making process [49–51], within the defense-related fusion community has also revitalized interest in DS theoretic (DST) methods [14, 17, 52, 53]. One reason for this is the close relationship between DS theory and the Bayesian probabilistic framework, which forms the backbone of most of the methods for multi-sensor hard data fusion. Moreover, DST models can easily capture data imperfections that are characteristic of soft evidence, such as “uncertain” implication rules which are very difficult to capture using, for example, the Bayesian probability framework [54, 55]. DS theory, given its flexibility in representing various data imperfections typical of soft data and its close relationship to existing multi-sensor fusion techniques that are used on probability theory, forms the ideal basis for development of a soft/hard fusion framework. Here, we present an analytical soft/hard data fusion framework addressing the challenges discussed in Section 1.3. This work consists of six main contributions.

1.4.1 Conditional Core Theorem

As in probability theory, conditioning is the primary tool for DST evidence updating [56]. In fact, the core of the proposed technique for soft/hard fusion is based on DST conditionals. Hence, it is imperative that we fully understand the various conditional notions and their implications. The conditional core theorem addresses this by providing a complete characterization of a DST conditional notion that is probably the most suitable for the task at hand. In addition to providing insight into the conditioning operations (and fusion operations based upon conditionals), one direct contribution of this important result is that it establishes the theoretical grounds for developing algorithms for efficient computation of DST conditionals. We also address the converse problem: what evidence may have played a role in generating a given change of belief (i.e., conditioned belief)? This converse result can be of significant practical value in certain applications (e.g., for studying the sensitivity of the updated

knowledge base with respect to the evidence received).

1.4.2 Conditional Approach to Evidence Fusion

We propose a new DST strategy for soft/hard data fusion that addresses the challenges \mathbf{C}_1 - \mathbf{C}_4 . The proposed strategy can handle sources possessing non-identical scopes with the least amount of computational burden compared to alternative methods that require the scopes to be expanded. In addition, it possesses several other intuitively appealing properties such as tolerance to contradictory evidence, ability to “refine” and “filter-out” information depending on the environment (e.g., prevailing threat level, etc.). It also satisfies most of the algebraic properties and characteristics that one expects fusion operators to possess. We also establish criteria for different fusion environments and propose various parameter selection strategies.

1.4.3 Consensus in DST Fusion Environments

Consensus is an important fusion problem. Here, we propose a consensus protocol applicable to soft/hard data environments. The proposed consensus protocol is defined in a setting that is applicable to any general fusion network (i.e., fully or partially connected, synchronous or asynchronous communications), albeit certain “coupling conditions” that are needed to guarantee the convergence (i.e., the existence of a consensus). As far as we know, the proposed method is probably the most general consensus protocol in the sense that **(i)** it applies to fully or partially connected networks with possible communication delays; and **(ii)** the consensus is sought among a set of DST data models, thus allowing one to analyze/generate consensus in a multitude of application domains that call for the use of richer and more flexible models to handle the underlying uncertainties (compared to simple opinion dynamic models and probabilistic models).

1.4.4 Credibility Estimation in the Absence of Ground Truth

Due to its subjective nature which can otherwise compromise the integrity of the fusion process, it is critical that soft data be validated prior to its incorporation into the fusion engine. The strategy of *discounting* evidence based on source reliability may not be applicable when dealing with soft sources because their reliability (e.g., an eye witnesses account) is often unknown beforehand.

We propose a methodology based on the notion of consensus that we have already developed to estimate the credibility of (soft) evidence in the absence of a “ground truth.” This estimated credibility can then be used for source reliability estimation, discounting or appropriately “weighting” evidence for fusion. The consensus procedure is set up via the proposed evidence fusion methods, thus allowing it the flexibility to capture a variety of imperfections inherent to soft evidence.

1.4.5 Efficient Computation of DST Conditionals

Even though the conditional core theorem provides a complete characterization of the conditional notions and in some cases can provide a significant reduction in computational overhead, the numerical calculation of the conditionals can still be computationally prohibitive. To address this, we propose a graphical procedure and an associated message passing scheme for efficient computation of the conditionals. The computational complexity of conditional computations is dependent on the evidence and how the evidence is modeled in addition to other algorithmic parameters. Thus, the extent of the computational gains can only be estimated after a careful analysis of the implementation and application details. Keeping these observations in mind, we provide criteria, which are based on “low cost” bounds, that one may use to determine if the use of the conditional core and the proposed method indeed offers computational gains.

1.4.6 Approximations Via Statistical Sampling

The computational overhead of DST methods can still be significantly high, especially in the soft/hard data domains, where one often has to deal with sources having large scopes. An efficient fusion framework must be capable of mitigating the computational burden via reasonable approximations, thus generating fast, yet sufficiently accurate results. We propose a method based on statistical sampling techniques to reduce the computational burden by approximating the DST data models by simpler and computationally efficient variants.

1.5 Organization of the Dissertation

The rest of the dissertation is organized into three main parts and an epilogue dedicated to the future research directions and concluding remarks.

Part I: Preliminary Notions contains introductory material that is essential to understanding the rest of the dissertation.

Chapter 2 provides a brief review of data fusion with an introduction to popular fusion models and levels of abstraction.

Chapter 3 presents a review of uncertainty processing, where we provide a taxonomy of uncertain data types and popular formalisms available for working with such data. We also provide a detailed review of DS belief theory along with the notions relevant to the work presented in this dissertation.

Part II: Soft/Hard Data Fusion Framework contains the proposed analytical framework.

Chapter 4 provides a brief review of DST conditional notions. A mathematical result that characterizes these conditional notions is then derived. The converse

problem is also addressed. Theoretical results are explained with numerical and application examples.

Chapter 5 introduces the main fusion component of the proposed framework. We discuss the issue of working with sources having non-identical scopes of expertise and then derive a new conditional notion that applies to both identical and non-identical scopes. Based on this convenient representation, we present a new fusion rule and derive its various properties. We illustrate several fusion strategies and also provide insight into parameter selection;. The proposed methods are then illustrated with examples; several guidelines are provided via a detailed illustrative example. Two real-life application examples are also provided to further illustrate the DST modeling and application of the fusion techniques.

Chapter 6 studies the problem of consensus generation. We provide an overview of the problem and challenges in generating a consensus in soft/hard data environment. The consensus generation is then established as a special case of convergence in non-linear asynchronous iterations. We provide a brief review of the theory of non-linear paracontractions and then set up a consensus generation protocol for soft/hard data modeled via DST notions. Consensus process and properties therein are illustrated with a detailed example.

Chapter 7 presents a consensus-based technique for estimation of credibility of evidence in the absence of ground truth. The proposed method is illustrated via a numerical example.

Part III: Computational Optimizations contains several techniques for efficient computation and reduction of computational burden associated with DST methods.

Chapter 8 presents an efficient method, based on graph theoretic and message passing techniques, for computation of DST conditionals. The proposed method is

based on the theorem that was developed in Chapter 4. Theoretical bounds for the resultant computational gains are derived, and they are evaluated and verified via Monte-Carlo simulations.

Chapter 9 presents an approximation technique based on statistical sampling techniques for reduction of computational burden associated with DST data representations. The proposed method is evaluated via Monte-Carlo techniques simulating typical real-life scenarios.

Part IV: Epilogue contains the concluding chapters of the dissertation.

Chapter 10 presents several future research directions to extend the proposed work and also to generate new research avenues based on the theories and techniques that have been developed here.

Chapter 11 provides concluding remarks.

Appendix contains detailed proofs of the various results that appear in Chapters 4, 5 and 6.

Vita contains the author's biography.

Part I

PRELIMINARY NOTIONS

Multi-sensor Data Fusion

DATA fusion, in a broad sense, involves combining information to better estimate the state of some aspect of universe. Multi-sensor data fusion (i.e., fusion of data from multiple sensors) is an emerging technology among both military and non-military (or civil) application domains. In this chapter, we provide a basic overview of various fusion notions and abstraction levels in order to lay a foundation for the rest of the presentation. A thorough understanding of the fundamental differences of fusion and knowledge abstraction levels is extremely helpful in understanding a given fusion problem as well as in choosing appropriate fusion method(s) for a given task.

2.1 Data Fusion

Data fusion is defined in the initial *Data Fusion Lexicon* produced by Joint Directors Laboratories Data Fusion Subgroup as,

a process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats, and their significance. The process is characterized by continuous refinements of its estimates and assessments, and the evaluation of the need for additional sources, or modification of the process itself, to achieve improved results [4].

Multi-sensor fusion refers to the case where data are fused from multiple sensors, whereas in the case of single-sensor fusion, multiple data from a single sensor are fused. Thomopoulos in [57] provides a discussion regarding the advantages of multiple-sensor systems over single-sensor systems. This discussion states that there are numerous advantages in using multiple sensor systems including:

- higher signal-to-noise ratio;
- increased robustness and reliability in the event of sensor failures;
- information regarding independent features in the system can be obtained;
- extended parameter coverage rendering a more complete picture of the system;
- increased dimensionality of the measurement;
- improved resolution, reduced uncertainty and increased confidence;
- increased hypothesis discrimination with the aid of more complete information arriving from multiple sensors;
- reduction in measurement time, and possibly costs - there is a trade off to consider in this issue. Thus, an optimal number of sensors to extract the required information from a system should be ideally pursued.

2.1.1 Processing Architectures

Three main fusion architectures can be identified as follows.

Data-level Fusion: if the sensors are measuring the same physical property (or properties), then the raw sensor measurements can be directly combined. Typical raw sensor fusion techniques involve classical estimation techniques, e.g., *Kalman Filtering*.

Feature-level Fusion: involves extraction of a representative set of features from sensor data and then combined into a single vector of features. Typical techniques include pattern recognition techniques such as neural networks, clustering algorithms, or template-based methods.

Decision-level fusion: Here, each sensor makes a preliminary estimation of an entity's identity in terms location, attributes, or any other DJDJDDJKDK. The fusion techniques at this level includes weighted decision methos (e.g., voting), classical inference, Bayesian inference, and Dempster-Shafer theoretic methods.

Remark: Mention that the fusion methods developed here applies to decision-level fusion, and can also be applied to feature-level fusion.

2.2 Fusion Models

Data fusion models are a result of attempts by many in developing a unified terminology. One of the first such attempts at unifying the terminology (still within military applications) resulted in, perhaps the most widely known, JDL process model, which was developed by the Joint Directors of Laboratories (JDL) data-fusion sub-panel, under US Department of Defense (DoD).

Level 1 Object Refinement: attempts to locate and identify objects. For this purpose a global picture of the situation is reported by fusing the attributes of an object from multiple sources. The steps included at this stage are: Data alignment, prediction of entitys attributes (i.e. position, speed, type of damage, alert status, etc.), association of data to entities, and refinement of entitys identity.

Level 2 Situation Assessment attempts to construct a picture from incomplete information provided by level 1, that is, to relate the reconstructed entity with an observed event (e.g. aircraft flying over hostile territory).

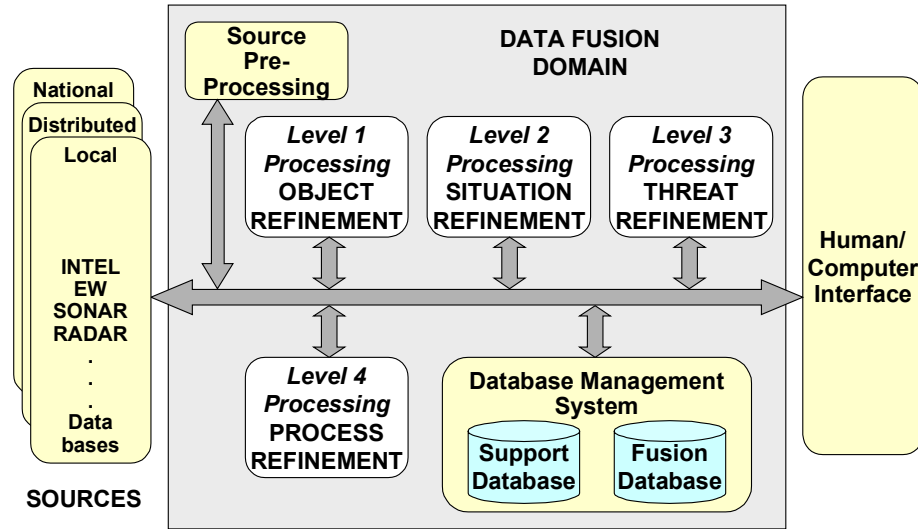


Figure 2.1: JDL Fusion Model (1992 Revision) [1]

Level 3 Threat Assessment interprets the results from level 2 in terms of the possible opportunities for operation. It analyses the advantages and disadvantages of taking one course of action over another.

Level 4 Process Refinement performs the main functions of (i) monitoring the data fusion process performance, (ii) identification of information needed to improve multilevel fusion product, (iii) determination of source specific requirements to collect relevant information and (iv) allocation of sources to mission goals.

2.3 Challenges for Multi-sensor Fusion

One needs to address several fundamental issues when building a data fusion system for a given application:

(i) what algorithms or techniques are appropriate and optimal for a particular application ?

(ii) what architecture should be used (i.e., where in the processing flow should data be fused) ?

(iii) how should the individual sensor data be processed to extract the maximum amount of information ?

(iv) what accuracy can realistically be achieved by a data fusion process?

(v) how can the fusion process be optimized in a dynamic sense?

(vi) how does the data collection environment (i.e., signal propagation, target characteristics, etc.) affect the processing;

(vii) under what conditions does multisensor data fusion improve system operation?

Answers to all these questions, help one to narrow down the available options and to understand the optimal fusion strategy as well as an appropriate uncertainty processing framework to be used.

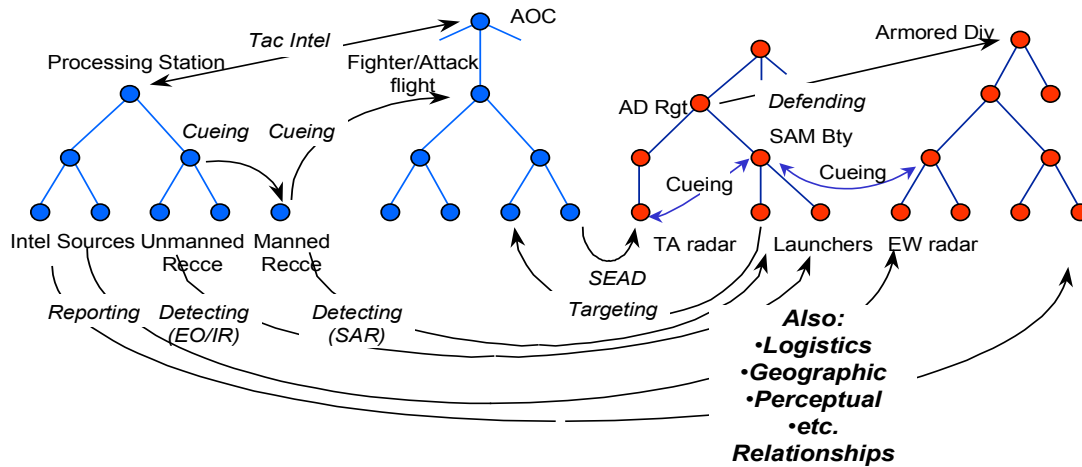


Figure 2.2: A Complex Fusion Scenario [1]

As we have already discussed, soft/hard data fusion is a special case of multi-sensor fusion, where data are fused from both human-based and physics-based sensors. Soft/hard data fusion is characterized by the numerous differences in scope of sources,

unknown reliability, unknown distributions, unstructured and often inconsistent data. These imperfections and differences in sources and characteristics of the data makes hard/soft fusion a much harder problem than regular multi-sensor fusion. Figure 2.3 illustrates a complex fusion scenario, working at several fusion levels, e.g., object refinement and situation assessment.

The other most important aspect of designing fusion frameworks is the identification of an appropriate uncertainty processing framework(s) to handle the uncertainties characteristic of the targeted application(s). In Chapter 3, we provide a discussion on various types of uncertainty and uncertainty processing formalisms that are popularly used in fusion applications working at various fusion levels.

Uncertain Data Processing

UNCERTAIN data processing lies in the heart of a data fusion framework. Fusion characteristics and performance of such frameworks are often dominated by the mechanism(s) used to handle uncertain data. Conversely, fusion mechanism design requires a proper understanding of **(i)** the types of uncertain data that needs to be processed, **(ii)** fusion and inferencing requirements, and **(iii)** computational constraints (e.g., memory, processing time, etc.) of the given task. A sound knowledge of the types of data uncertainties, processing mechanisms and their limitations is of critical importance in scenarios that call for data fusion. Therefore, we provide a detailed discussion on uncertain data types, their origins and three uncertainty processing formalisms that are popular among the fusion community. In addition, we provide a detailed summary of basic notions of DS belief theory.

This Chapter is organized as follows: Section 3.1 discusses how different types of uncertain data are generated and presents a classification for uncertain data; Section 3.2 provides a brief summary of several approaches, which are popular among fusion researchers, for handling different types of uncertain data; Section 3.3 provides a comparison of these uncertainty processing formalisms; Section 3.4 provides a detailed summary of DS belief theory, the uncertainty processing formalism used in the

work presented here; and Section 3.5 contains the chapter summary.

3.1 A Taxonomy of Uncertain Data

Uncertain data arises as a result of lack of information (regarding a subject of interest). The type of information that is lacking consequently gives rise to different types of uncertainty, which can be broadly classified as follows [58–60]⁴:

Incompleteness: refers to the case where certain parts (or all) of the information are missing from a database record. For example, consider a database record `DB_REC` containing the information of a detected unit in a battlefield as `DB_REC={UNIT_TYPE, LOCATION, SPEED}`. Now, for instance, if the `UNIT_TYPE` is missing from the database record, then `DB_REC` is *incomplete*.

Imprecision: refers to the case where the given information is not having the required precision. For instance, in the above battlefield scenario, if the `SPEED` of an enemy unit has been entered as `25 – 40 mph`, it may not be precise enough, for instance, to fire an unguided rocket. In this case, `DB_REC` is *imprecise*.

Uncertain: refers to the case where the given information is precise, complete, but *uncertain* since it may be wrong. This type of uncertainty occurs mainly due to imperfections associated with the source of information. For example, in the above battlefield scenario, suppose the soldier in charge, due to bad weather conditions, could only identify the `UNIT_TYPE` of a detected unit to be either a `TANK` or `JEEP` with 75% confidence. In this case, the information about the detected unit becomes uncertain, even though the database record can still be complete and precise.

In many applications, data records are often constrained by the domain and range restrictions for categorical and numerical data types, respectively. For instance, in the

⁴Authors identify incompleteness, imprecision and uncertainty as types of *ignorance*, which is defined as lack of information and is eventually related to uncertainty.

above battlefield scenario, it is known that the `UNIT_TYPE` can only be, for instance, `TANK`, `JEEP` or `TRUCK`, and the `SPEED` must lie in the interval $[0, 250]$ *mph*. In this case, not being able to identify the `UNIT_TYPE` can actually be viewed as not having enough precision to differentiate between `TANK`, `JEEP` or `TRUCK`. This justifies the view where incompleteness is taken as a special case of imprecision. Moreover, incompleteness and imprecision are context-dependent in the sense that information that is imprecise or incomplete in one context may be complete and precise (at least, precise enough) in a another. For example, detecting the `SPEED` of enemy unit as $25 - 40$ *mph* may not be precise enough to fire an unguided rocket, but it is definitely precise enough to issue an alert of a possible intrusion. Following this line of argument, one can clearly see that all of the above cases are different types of uncertainty resulting from lack of information.

Type	Subtype	Example
Incompleteness	Existential	An enemy unit has been detected at <code>LOCATION=LOC#1</code> , but <code>TYPE</code> is unknown
	Universal	All units that have been detected so far at various locations are hostile, but their <code>TYPE</code> is unknown
Imprecision	Disjunctive	The <code>UNIT_TYPE</code> of an enemy unit is detected as <code>TANK</code> or <code>JEEP</code>
	Negation	The <code>UNIT_TYPE</code> is not <code>TANK</code> .
	Interval-valued	The <code>SPEED</code> of a unit is detected as $25 - 40$ <i>mph</i> .
	Fuzzy-valued	A tracked unit is moving <i>very fast</i> .
Uncertainty	Probability	The chance of <code>UNIT_TYPE=TANK</code> is 70%.
	Possibility	It is possible that the <code>SPEED=30 mph</code> for the detected unit.
	Credibility	My degree of belief for <code>UNIT_TYPE=TANK</code> is 0.8.

Table 3.1: A Classification of Uncertain Data Types

Table 3.1 presents a summary of different types and sub-types of uncertainties [60]. Note how the way evidence is captured/generated naturally leads to different types of uncertainty. Let us now analyze the mechanisms available for working with such uncertain data.

3.2 Approaches to Handling Uncertainty

A typical data fusion engine employs one or more uncertainty handling mechanisms for data representation, fusion, inferencing, decision-making and various other fusion related tasks. Let us proceed by formally defining the *problem of uncertainty handling*, as applicable for the task of data fusion in uncertain data domains.

Suppose a decision-maker (e.g., an analyst or a base commander) has the universe of discourse $\Omega = \{\omega_1, \dots, \omega_n\}$ ⁵ at hand as the set of likely hypothesis for a given decision-making problem. These hypothesis need to be associated with some values, be they probabilities, possibility measures, or any other quantitative terms suitably defined by the decision-maker, representing the decision-maker's quantified degrees of belief in their likely occurrences based on the available uncertain evidence [60].

Depending on the complexity and flexibility of the mechanism used, the types of uncertainty that can be represented and also the types of inferencing provided by a given formalism varies. For example, in the Bayesian probability theory, one assigns values to elements to a σ – *algebra* of Ω —which identifies a set of outcomes—to represent the degree of belief (e.g., of an analyst) about how likely the events (as defined by the σ – *algebra*) are to occur based on the evidence. Even though the probability theory has been the predominant paradigm for handling uncertainty over several decades, due to the way evidence is modeled, it is not capable of representing, for instance, a *disjunction* of two events defined in the σ – *algebra* of Ω (see Table 3.1) without making simplifying assumptions. In his seminal work [61], Pearl identifies four main classes of approaches to uncertainty handling, namely, *neo-probabilist*, *neo-calculist*, *neo-logicist*, and *neo-possibilist*, which are motivated by these weaknesses in traditional frameworks, such as probability theory and classical logic. Table 3.2 [60] provides a summary of these approaches.

⁵The symbol Ω is most commonly used in probability theory to denote the universe of discourse or the **sample space**. However, in DS theoretic literature sample space is referred to as **frame of discernment** and the symbol Θ is used instead of Ω . See Section 3.4 for a detailed discussion.

Category	Motivation	Definition	Technologies
Neo-probabilist	Traditional probability framework is not suitable for handling large number of variables	Consistent with the traditional probabilistic framework; efficient computational methods are provided for various fusion tasks	Bayesian networks
Neo-calculist	Probability calculus is inadequate for capturing uncertainty	New calculus for handling uncertainty and fusion operations	DS theory; certainty factors
Neo-logicist	Monotonicity in classical logics is not suitable for handling commonsense knowledge	Deals with non-numerical and non-monotonic inferencing for handling uncertainty in negation	Default logic, program completion
Neo-possibilist	Coarse grained two-valued truth representation is not sufficient to model practical applications	Nothing is a matter of degree in two-valued Boolean logic	Fuzzy sets, fuzzy logic, possibility theory, possibilistic logic

Table 3.2: Approaches to Handling Uncertainty in Data

Let us now briefly analyze the basic differences of several popular uncertainty modeling formalisms, namely, Bayesian probability theory, possibility theory, and DS theory, in order to understand how these fundamental differences affect the types of uncertainty that can be accommodated.

3.2.1 Bayesian Probability Theory

A Bayesian probabilistic model consists of the triplet (Ω, Φ, P) , where Ω is the *sample space*, Φ is a σ -algebra of subsets of Ω , and P is a non-negative mapping of Φ into the interval $[0, 1]$. The probability assignment $P : \Phi \mapsto [0, 1]$ possesses the following properties:

Axiom 1: $\Omega \in \Phi$ with $P(\Omega) = 1$;

Axiom 2: If $A \in \Phi$, then $\bar{A} \in \Phi$, where $\bar{A} = \Omega \setminus A$;

Axiom 3: For pairwise disjoint sets A_n for $n \geq 1$,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Bayes' theorem is employed to compute the probability of a hypothesis, given some observation event. Consider a collection of hypotheses $\{H_i\}_{i=1}^n$, $H_i \in \mathcal{F}$, for $i = 1, \dots, n$, and suppose $P(E) > 0$ for some event E (typically, E is some evidence). *Conditional probability* of occurring H_i provided the evidence E is given by,

$$P(H_i|E) = \frac{P(E|H_i) P(H_i)}{P(E)}, \quad (3.1)$$

where

$$P(E) = \sum_{j=1}^n P(E|H_j) P(H_j) \quad (3.2)$$

The quantities $P(H_i)$ and $P(E|H_i)$ are termed *a-priori probabilities*, since they represent statements that can be made prior to knowing the true subject of any observation. The conditional probabilities are calculated using above quantities along with the probability of observation event $P(E)$. These conditional probabilities are then combined using the *generalized Bayesian inference formula* [9]:

$$P(H_i|E_1 \cap \dots \cap E_k) = \frac{P(H_i) \cdot P(E_1|H_i) \dots P(E_k|H_i)}{\sum_{j=1}^n P(H_j) \cdot [P(E_1|H_j) \dots P(E_k|H_j)]} \quad (3.3)$$

to obtain *a-posteriori probabilities* with respect to the totality of events in the whole sample space. A suitable decision logic is utilized to arrive at a decision based on these final probabilities. *Maximum a-posteriori (MAP)* and *maximum likelihood (ML)* methods are widely used as Bayesian decision rules [62, 63].

Each proposition $A \in \mathcal{F}$ is associated with the probability $P(A)$. Thus, from Axioms 1-3, we must have

$$P(A) + P(\overline{A}) = 1, \quad (3.4)$$

for all $A \in \mathcal{F}$. Hence, in a Bayesian model, the knowledge one has about a proposition A determines explicitly the knowledge one has regarding its complement. We cannot refrain from assigning probability numbers to events in \overline{A} that we are not certain of. Therefore, it is incapable of representing the ignorance we may have regarding the events in \overline{A} . Moreover, the difficulties in defining prior likelihoods when sufficient information is not available, and the requirement that the competing hypotheses must be mutually exclusive are other disadvantages of Bayesian probability theory [9].

3.2.2 Possibility Theory

Possibility measures introduced by Zadah [64] is closely associated with *fuzzy sets and measures* [65]. It considers a body of knowledge represented as subsets of a reference set \mathcal{S} . A set of functions $C : \Theta \mapsto [0, 1]$ referred to as *confidence functions* map the elements of Θ , which is defined as the powerset of \mathcal{S} , i.e., $\Theta \equiv 2^{\mathcal{S}}$. These confidence functions are *monotonic*, i.e., if $A \subseteq B \subseteq \Theta$, then $C(A) \leq C(B)$. The interpretation of this property is that, if an event A implies a second event B , then there is at least as much confidence in the occurrence of the event B as in the occurrence of the event A . The consequences of this monotonic property are:

$$C(A \cup B) \geq \max(C(A), C(B)); \quad \text{and} \quad (3.5)$$

$$C(A \cap B) \leq \min(C(A), C(B)). \quad (3.6)$$

The limiting case $C(A \cup B) = \max(C(A), C(B))$ defines what are referred to as *possibility measures* [65]. Suppose $E \in \Theta$ is such that $C(E) = 1$. Possibility measure Π_{pos} is defined as $\Pi_{pos}(A) = 1$ if $A \cap E \neq \emptyset$ and 0 otherwise. We interpret $\Pi_{pos}(A) = 1$ as A is possible. Also note that $\Pi_{pos}(A \cup \overline{A}) = \Pi_{pos}(\mathcal{S}) = 1$ and $\max(\Pi_{pos}(A), \Pi_{pos}(\overline{A})) = 1$. Using this, A and \overline{A} can be interpreted as two contradictory events, i.e., at least one event is possible. However, one event being possible does not prevent the other being possible as well. The notion of $C(A \cup B) = \max(C(A), C(B))$ seem to be consistent with possibility in the real world, i.e., oc-

currence of $A \cup B$ requires only the easiest event (most possible event) of the two to happen. Similarly, using the other limiting case $C(A \cap B) = \min(C(A), C(B))$, a *necessity measure* \mathcal{N}_{pos} is defined to interpret that if an event is necessary, its contrary is impossible. Conversely, if an event is possible, its contrary is absolutely not necessary.

Uncertainty of events can be characterized by these possibility and necessity functions, thus weakening the *additivity property* (3.4) of the Bayesian framework. Fuzzy membership functions defining various relationships among fuzzy sets serve as possibility distribution functions. Approaches based on fuzzy reasoning can represent the vagueness of information. Probability theory only allows us to represent the chance of extremes (occurrence or non-occurrence) of an event while possibility theory could extend our view over to notions of “to what extent the event is possible?” and “to what extent the event is necessary?” In certain situations, where such vagueness of information needs to be represented, this formalism offers advantages over probability theory [66]. The disadvantages of this framework include increased number of computations compared to other methods, and the potential difficulty in generating suitable membership functions corresponding to the fuzzy sets.

3.2.3 DS Theory

The DS belief theory, originally proposed by Dempster, can be thought of as an extension to probability theory in the following sense. In comparison to the triplet (Θ, \mathcal{F}, P) in probability theory, the triplet $(\Theta, \mathfrak{F}_\Theta, m_\Theta(\cdot))$ defines a DST model of data, where Θ is referred to as the *frame of discernment (FoD)*. The function $m_\Theta : 2^\Theta \mapsto [0, 1]$ is a mapping that assigns non-negative support to subsets of Θ and \mathfrak{F}_Θ is the *Core* which contains subsets B of Θ s.t. $m_\Theta(B) > 0$. Propositions $B \subseteq \Theta$ for which $m_\Theta(B) > 0$ are referred to as *focal elements*. The mapping $m_\Theta(\cdot)$, which is

referred to as *basic probability assignment* or *mass function* or *mass* satisfies:

$$m_{\Theta}(\emptyset) = 0; \quad \text{and} \quad \sum_{B \subseteq \Theta} m_{\Theta}(B) = 1. \quad (3.7)$$

The *belief* and *plausibility* functions in DS theory represents the total support that can move into a proposition without any ambiguity and the extent to which one finds a proposition plausible, respectively. When the focal elements consists of a single element in Θ , i.e., $|B| = 1$ for all $B \in \mathfrak{F}$, both belief and plausibility functions reduce to probability mass functions. Often, the belief and plausibility functions are seen as lower and upper envelopes of the family of all probability distributions conforming to DS model (based on evidence). Hence, one may draw a comparison between the necessity and possibility measures in possibility theory to belief and plausibility measures in DS theory.

Data are often modeled in terms of mass functions and probability mass functions in DS theory and probability theory, respectively. A mass functions assigns support to all subsets of Θ . A probability mass function P only assigns support to elements of a σ – *algebra* of Θ defining the measurable events of Θ under P ; further, axioms 1-3 of governing P results in $P(\overline{B}) = 1 - P(B)$ for any outcome B in the σ – *algebra*; thus, support of any outcome also defines the support of its complement, which may become very restrictive in certain application where available evidence may not provide any information on its complement. The higher modeling flexibility of DS theory compared to probability theory is a result of these facts. DS literature also provides an extensive array of methods for various data fusion and decision-making tasks. The disadvantage of complexity and additional modeling flexibility is the well-known exponentially large computational overhead of DST methods.

3.3 A Comparison of Popular Formalisms

Consider the battlefield scenario, where a detected hostile unit is to be modeled as $\text{DB_REC} = \{\text{UNIT_TYPE}, \text{LOCATION}, \text{SPEED}\}$ based on the available (uncertain) evi-

dence. The type of uncertainties that can be captured is dependent on the formalism used. Moreover, the fusion output which is to be used for decision-making is also dependent on this choice.

To illustrate, let us consider a fusion result on the attribute `UNIT_TYPE` based on some typical uncertain evidence. In probability framework, evidence must be distributed and assigned to all elements of `UNIT_TYPE`, i.e., `TANK`, `JEEP` and `TRUCK`. For instance, if one obtained evidence, say on `UNIT_TYPE = JEEP \cup TRUCK` with certain confidence, this support has to be distributed among the individual elements `JEEP` and `TRUCK`. The fusion results on this will also be a discrete (continuous if a attribute is continuous) probability distribution on all these individual elements (see Table 3.3). However, in the case of DST, one can assign support to composite elements. For example, the proposition `JEEP \cup TRUCK` can be assigned a support without distributing into individual elements. The fusion results, depending on the evidence and fusion mechanisms used, can consist of composite propositions. Table 3.3 inspired by [60]) provides a summary of output uncertainty provided by different formalisms.

Formalism	Output Uncertainty	Example
Bayesian probability	Discrete or continuous distribution	TN=0.1, JP=0.2, TR=0.7
DS theory	Mass distribution among focal elements	TN=0.1, (JP, TR)=0.3, JP=0.6
Certainty factors	Certainty factor values	TN= - 0.2, JP=0.2, TR=0.6
Classical logic	Disjunction	TN \cup JP
Possibility	Possibility measure	TN=0.3, JP=0.6, TR=0.4

Table 3.3: Uncertainty output types from different formalisms for a typical fusion result of uncertain evidence on `UNIT_TYPE`, where TN=TANK, JP=JEEP and TR=TRUCK.

3.4 Detailed Review of DS Theory

As we have already discussed, DS belief theory provides a convenient framework for representing and working with a wide variety of data imperfections. Over the years,

DS belief theory has emerged as one of the most dominant frameworks for uncertainty processing for decision-making purposes in a wide spectrum of problem domains. Here, we provide a basic review of the basics of DST that forms the foundation of the work provided in the chapters to follow. Some of the notions, terms, and abbreviations have already been introduced and discussed in the previous chapters. However, in this section, we provide a comprehensive review (which may involve redefinition of relevant notions, terms, and abbreviations) for the sake of completeness. The notation and terminology introduced here will be used throughout the rest of the presentation.

3.4.1 Basic Notions

In DST, the total set of mutually exclusive and exhaustive propositions that a node may discern is referred to as its FoD (Frame of Discernment) $\Theta = \{\theta_1, \dots, \theta_n\}$; it signifies the ‘scope’ of expertise. A proposition θ_i represents the lowest level of discernible information; it is referred to as a *singleton*. We use $|\Theta|$ and 2^Θ to denote the cardinality and the power set of Θ , respectively. Elements in 2^Θ form all the propositions of interest in DST. We use $\Theta \setminus B$, or simply \overline{B} when the FoD is clear from the context, to denote all singletons in Θ that are not included in B . In DST, the ‘support’ for proposition B is provided via a *basic belief assignment (BBA)* or *mass assignment* or *mass function*:

Definition 1 (Mass Function or BBA). *The mapping $m_\Theta : 2^\Theta \mapsto [0, 1]$ is a mass function for the FoD Θ if $m_\Theta(\emptyset) = 0$ and $\sum_{A \subseteq \Theta} m_\Theta(A) = 1$. \square*

The mass assigned to a proposition is free to move into the individual singletons that constitute the *composite proposition*⁶, thus generating the notion of *ignorance*. The propositions that possess nonzero mass form the *core* \mathfrak{F}_Θ . The triplet $\{\Theta, \mathfrak{F}_\Theta, m_\Theta(\cdot)\}$, denoted by \mathcal{E}_Θ , is the corresponding *body of evidence (BoE)*. The

⁶We will use the term, composite proposition to denote a proposition which consists of more than one singleton, i.e., to denote a proposition $B \subseteq \Theta$ s.t. $|B| > 1$.

quantity $m_\Theta(B)$ measures the support assigned to proposition $B \subseteq \Theta$ *only*. The belief assigned to B on the other hand takes into account the supports for all proper subsets of B as well:

Definition 2 (Belief and Plausibility). *Given a BoE $\{\Theta, \mathfrak{F}_\Theta, m_\Theta(\cdot)\}$ and a proposition $B \subseteq \Theta$, $\text{Bl}_\Theta : 2^\Theta \mapsto [0, 1]$, where $\text{Bl}_\Theta(B) = \sum_{C \subseteq B} m_\Theta(C)$ is the belief of B ; and $\text{Pl}_\Theta : 2^\Theta \mapsto [0, 1]$, where $\text{Pl}_\Theta(B) = 1 - \text{Bl}_\Theta(\overline{B})$ is the plausibility of B . \square*

In other words, in DST, $\text{Bl}_\Theta(B)$ represents the total support that can move into the proposition B without any ambiguity, and $\text{Pl}_\Theta(B)$ represents the extent to which one finds the proposition B plausible. We use $\hat{\mathfrak{F}}_\Theta$ to denote the set of propositions with non-zero belief, .viz, $\hat{\mathfrak{F}}_\Theta \equiv \{B \subseteq \Theta \mid \text{Bl}_\Theta(B) > 0\}$. When all the focal elements are singletons, i.e., $|B| = 1$ for all $B \in \hat{\mathfrak{F}}_\Theta$, the mass, belief and plausibility functions reduce to probability mass functions. A probability distribution $\text{Pr}_\Theta(\cdot)$ such that $\text{Bl}_\Theta(B) \leq \text{Pr}_\Theta(B) \leq \text{Pl}_\Theta(B)$, $\forall B \subseteq \Theta$, is said to be *compatible* with the underlying mass function $m_\Theta(\cdot)$. An example of such a probability distribution is the *pignistic probability distribution* $\text{BetP}_\Theta(\cdot)$ [67]

$$\text{BetP}_\Theta(\theta_i) = \sum_{\theta_i \in B \subseteq \Theta} \frac{m_\Theta(B)}{|B|}. \quad (3.8)$$

3.4.2 Conditional Notions

As in probability theory, *conditioning* is the primary tool for DST evidence updating [56]. For a BoE $\mathcal{E}_\Theta \equiv \{\Theta, \mathfrak{F}_\Theta, m_\Theta(\cdot)\}$ and a conditioning event $A \subseteq \Theta$, the *conditional mass, belief and plausibility* functions are denoted as $m_\Theta(\cdot|A)$, $\text{Bl}_\Theta(\cdot|A)$ and $\text{Pl}_\Theta(\cdot|A)$, respectively. Dempster in [68] proposed:

Theorem 1 (Dempster's (DS) Conditionals). *For any $B \subseteq \Theta$, the conditional belief $\text{Bl}_\Theta(B|A) : 2^\Theta \mapsto [0, 1]$ is given by*

$$\text{Bl}_\Theta(B|A) = \frac{\text{Bl}_\Theta(B \cup \overline{A}) - \text{Bl}_\Theta(\overline{A})}{1 - \text{Bl}_\Theta(\overline{A})}, \quad (3.9)$$

where the conditioning event A satisfies $\text{Bl}_\Theta(\overline{A}) < 1$. ■

The work of Fagin, et al., in [69] proposes an alternative conditional notion⁷ based on the premise that DS belief and plausibility functions can be understood as inner and outer measures of non-measurable events, respectively.

Theorem 2 (Fagin-Halpern (FH) Conditionals). *[69] For any $B \subseteq \Theta$, the conditional belief $\text{Bl}_\Theta(B|A) : 2^\Theta \mapsto [0, 1]$ and conditional plausibility $\text{Pl}_\Theta(B|A) : 2^\Theta \mapsto [0, 1]$ are given by*

$$\text{Bl}_\Theta(B|A) = \frac{\text{Bl}_\Theta(A \cap B)}{\text{Bl}_\Theta(A \cap B) + \text{Pl}_\Theta(A \cap \overline{B})}; \quad (3.10)$$

$$\text{Pl}_\Theta(B|A) = \frac{\text{Pl}_\Theta(A \cap B)}{[\text{Pl}_\Theta(A \cap B) + \text{Bl}_\Theta(A \cap \overline{B})]}, \quad (3.11)$$

respectively, where the conditioning event A satisfies $\text{Bl}_\Theta(A) > 0$. ■

A proposition with positive mass after conditioning is referred to as a *conditional focal element*. The set of conditional focal elements is referred to as the *conditional core* and denoted by $\mathfrak{F}_{\Theta|A}$. Thus, $\mathfrak{F}_{\Theta|A} \equiv \{B \subseteq \Theta \mid m_\Theta(B|A) > 0, A \in \hat{\mathfrak{F}}_\Theta\}$, where $m_\Theta(\cdot|A) : 2^\Theta \mapsto [0, 1]$ is the corresponding conditional mass function related to $\text{Bl}_\Theta(\cdot|A)$ via the Möbius transform [28]:

$$m_\Theta(B|A) = \sum_{C \subseteq B} (-1)^{|B-C|} \text{Bl}_\Theta(C|A), \quad \forall B \subseteq \Theta. \quad (3.12)$$

3.4.3 Evidence Combination

Evidence combination is the process of combining BoEs $\mathcal{E}_{\Theta_i} \equiv \{\Theta_i, \mathfrak{F}_{\Theta_i}, m_{\Theta_i}(\cdot)\}$, for $i = 1, \dots, n$, to arrive at a new BoE $\mathcal{E}_\Theta \equiv \{\Theta, \mathfrak{F}_\Theta, m_\Theta(\cdot)\}$; we denote this process as $\mathcal{E}_\Theta \equiv \mathcal{E}_{\Theta_1} \bowtie \dots \bowtie \mathcal{E}_{\Theta_n}$. A strategy used for such combinations is referred to as a

⁷The work reported here is based on the FH conditionals due to several reasons. See Chapter 5 for a detailed discussion on various DST conditionals and suitability of FH conditionals for the task of soft and hard data fusion.

combination rule. These combination rules can be characterized by how masses are combined and how conflicting masses are redistributed. However, the majority of popular BPA-based combination rules are based on the following two basic components ($n = 2$ case):

Definition 3 (Conjunctive and Disjunctive Forms). *The conjunctive and disjunctive forms of combination are given by*

$$m_{\Theta_1 \cap \Theta_2}(B) = \sum_{C \cap D = B} m_{\Theta_1}(C) m_{\Theta_2}(D);$$

$$m_{\Theta_1 \cup \Theta_2}(B) = \sum_{C \cup D = B} m_{\Theta_1}(C) m_{\Theta_2}(D),$$

respectively, where $m_{\Theta_1 \cap \Theta_2}(\cdot)$ and $m_{\Theta_1 \cup \Theta_2}(\cdot)$ are the fused BBA in each case. The parameter $\mathcal{K} = m_{\Theta_1 \cap \Theta_2}(\emptyset)$ is referred to as the conflict. \square

The most commonly used combination rule in DS theory is the *Dempster's Combination Rule (DCR)* [70].

Definition 4 (Dempster's Combination Rule (DCR)). *The fused BBA generated by the DCR is denoted as \oplus and is given by*

$$m_{\Theta}(B) = \frac{1}{1 - \mathcal{K}} m_{\Theta_1 \cap \Theta_2}(B), \quad B \neq \emptyset,$$

whenever $\mathcal{K} = m_{\Theta_1 \cap \Theta_2}(\emptyset) \neq 1$. \square

Remarks:

- (i) The DCR requires $\Theta_i = \Theta$, for $i = 1, 2$, for combination. When $\Theta_1 \neq \Theta_2$, the two FoDs are extended to a common FoD Θ , via the *ballooning extension*.
- (ii) DCR is not defined for $\mathcal{K} = 1$.

The main drawback of the DCR is the counter-intuitive results it generates in the presence of conflicting evidence. Alternative rules overcome this issue by redistributing the conflicting mass (or the conflict given by \mathcal{K}) in numerous ways.

Definition 5 (Yager's Method (YGR)). [71] *The fused BBA generated by the YGR is*

$$m_{\Theta}(B) = \begin{cases} m_{\Theta_1 \cap \Theta_2}(B), & \text{for } B \neq \emptyset; \\ m_{\Theta_1 \cap \Theta_2}(B) + \mathcal{K}, & \text{for } B = \Theta. \end{cases}$$

□

Dubois and Prade proposed to distribute conflict among propositions that actually contribute to the conflict.

Definition 6 (DP1). *The fused BBA generated by the DP1 is*

$$m_{\Theta}(B) = m_{\Theta_1 \cap \Theta_2}(B) + \sum_{\substack{C \cup D = B; \\ C \cap D = \emptyset}} m_{\Theta_1}(C)m_{\Theta_2}(D), \quad B \neq \emptyset.$$

□

The adaptive rule proposed by Dubois and Prade for possibility theory framework automatically accounts for relative reliability of sources being combined. The transformation of this rule to DST yields another combination rule.

Definition 7 (DP2). [3] *The fused BBA generated by the DP2 is*

$$m_{\Theta}(B) = \frac{1}{\mathfrak{K}} \max \left[\frac{m_{\Theta_1 \cap \Theta_2}(B)}{1 - \mathcal{K}}, \min(\mathcal{K}, m_{\Theta_1 \cup \Theta_2}(B)) \right], \quad B \neq \emptyset,$$

where

$$\mathfrak{K} = \sum_{B \subseteq \Theta} \max \left[\frac{m_{\Theta_1 \cap \Theta_2}(B)}{1 - \mathcal{K}}, \min(\mathcal{K}, m_{\Theta_1 \cup \Theta_2}(B)) \right].$$

□

Several versions of *partial conflict redistribution (PCR)* rules have been developed by Smarandache and Dezert [72], where the idea is to distribute the partial conflicts among the focal elements involved in the conflict. PCR5, hereinafter referred to as the PCR, is claimed to be the best among all the PCR rules.

Definition 8 (PCR). *The fused BBA generated by the PCR is*

$$m_{\Theta}(B) = m_{\Theta_1 \cap \Theta_2}(B) + \sum_{\substack{B \cap C = \emptyset; \\ C \subseteq \Theta}} \Gamma(B, C), \quad B \neq \emptyset,$$

where

$$\Gamma(B, C) = \frac{m_{\Theta_1}(B)^2 m_{\Theta_2}(C)}{m_{\Theta_1}(B) + m_{\Theta_2}(C)} + \frac{m_{\Theta_2}(B)^2 m_{\Theta_1}(C)}{m_{\Theta_2}(B) + m_{\Theta_1}(C)}.$$

□

A set of *robust combinations rules* (RCRs) proposed by Florea, et al., possesses self-adaptive behavior for conflicting evidence as well as non-identical frames. These are very important properties for a combination rule, especially in the case of hard/soft data fusion.

Definition 9 (RCR). *The fused BBA generated by the RCR is*

$$m_{\Theta}(B) = \alpha_{RCR}(\mathcal{K}) m_{\Theta_1 \cup \Theta_2}(B) + \beta_{RCR}(\mathcal{K}) m_{\Theta_1 \cap \Theta_2}(B), \quad B \neq \emptyset,$$

where $\alpha_{RCR}(\mathcal{K}) = 1 - (1 - \mathcal{K}) \beta_{RCR}(\mathcal{K})$; $\alpha_{RCR}(\cdot)$ and $\beta_{RCR}(\cdot)$ are increasing and decreasing functions respectively satisfying $\alpha_{RCR}(0) = 0$, $\alpha_{RCR}(1) = 1$, $\beta_{RCR}(0) = 1$ and $\beta_{RCR}(1) = 0$. □

3.4.4 Evidence Updating

Evidence updating is a special case of evidence combination, where, without loss of generality, a BoE $\mathcal{E}_{\Theta_1}[k] \equiv \{\Theta_1, \mathfrak{F}_{\Theta_1}[k], m_{\Theta_1}(\cdot)[k]\}$ is combined with BoEs $\mathcal{E}_{\Theta_i}[k] \equiv \{\Theta_i, \mathfrak{F}_{\Theta_i}[k], m_{\Theta_i}(\cdot)[k]\}$, for $i = 1, \dots, n$ to obtain an *updated* $\mathcal{E}_{\Theta_1}[k]$ given by $\mathcal{E}_{\Theta_1}[k+1]$. Here, k denotes discrete time/event index. We denote the updating process as $\mathcal{E}_{\Theta_1}[k+1] = \mathcal{E}_{\Theta_1}[k] \triangleleft \mathcal{E}_{\Theta_2}[k] \bowtie \dots \bowtie \mathcal{E}_{\Theta_n}[k]$, $\forall k$, (the order is as explained, unless parentheses are used).

3.4.5 Evidence Fusion: Non-exhaustive FoDs Case

The requirement to have identical FoDs in the BoEs being combined/updated constitutes a major drawback in some of the most widely used combination rules (e.g., the DCR). The approaches taken by fusion operators that can handle non-identical FoDs (so that $\Theta_1 \neq \Theta_2$ and $\Theta_1 \cap \Theta_2 \neq \emptyset$, for $n = 2$ case) can be categorized as follows.

3.4.5.1 Ignoring Differences in the FoDs

Each source would simply allocate zero mass to propositions that are not within its own FoD and continue applying the fusion operator. In essence, this approach assumes that each source can discern $\Theta_1 \cup \Theta_2$ and ignores the fact that some propositions are not within its scope of expertise. The counter-intuitive conclusions this approach may generate are well documented [73].

3.4.5.2 Deconditioning Approach

Here, each source would artificially introduce ambiguities into its evidence so that its own FoD is ‘deconditioned’ or ‘expanded’ to $\Theta_1 \cup \Theta_2$. For example, consider the *plausibilities correction method (PCM)* in [73] and let $\Theta_C = \Theta_1 \cap \Theta_2$. Then the propositions of $\Theta_1 \cap \overline{\Theta}_C$ are discerned by the first source alone, those of $\Theta_2 \cap \overline{\Theta}_C$ are discerned by the second source alone, and the propositions of Θ_C are discerned by both sources. A deconditioning step is applied to the partial knowledge of sources to ‘refer’ their knowledge to $\Theta_1 \cup \Theta_2$. Combination is performed via the multiplication of these ‘deconditioned’ plausibilities. The PCM requires that the plausibilities of only singleton propositions are maintained throughout. Due to the lack of non-singleton plausibilities after combination, it is impossible to obtain a valid BPA and the uncertainty intervals for any composite proposition.

3.4.5.3 Conditional Approach

The work in [74] proposes an evidence updating strategy that is based on FH conditionals and avoids the need to continually “expand” its FoD or introduce ambiguities when confronted with a source having a different FoD. Here, Belief update $\mathcal{E}_{\Theta_1}[k+1] = \mathcal{E}_{\Theta_1}[k] \triangleleft \mathcal{E}_{\Theta_2}[k]$ is given by [74]

$$\begin{aligned} \text{Bl}_{\Theta_1}(B)[k+1] = & \alpha(A)[k] \text{Bl}_{\Theta_1}(B)[k] \\ & + \beta(A)[k] \left[\text{Bl}_{\Theta_2}(B|A)[k] \right. \\ & \left. + \frac{1}{2} \sum_{\substack{\emptyset \neq X \subseteq B \cap A; \\ \emptyset \neq Y \subseteq A \setminus \Theta_1}} m_{\Theta_2}(X \cup Y|A)[k] \right]. \end{aligned} \quad (3.13)$$

Notice the following.

1. The term $\text{Bl}_{\Theta_1}(B)[k]$ weighted by $\alpha(A)[k]$ accounts for the evidence that is already available in $\mathcal{E}_{\Theta_1}[k]$ towards B .
2. The contribution from the second source (i.e., $\mathcal{E}_{\Theta_2}[k]$) consists of:
 - (a) the term $\text{Bl}_{\Theta_2}(B|A)[k]$ that accounts for evidence provided by those propositions that both Θ_1 and Θ_2 can discern (In fact, $\text{Bl}_{\Theta_2}(B|A)[k]$ computes the belief one allocates towards B when event A has occurred);
 - (b) the terms $m_{\Theta_2}(X \cup Y|A)[k]$ that account for the propositions that can only be discerned by Θ_2 , but can possibly move into Θ_1 later on (see Fig. 3.1). This type of evidence is generated by focal elements in $\mathcal{E}_{\Theta_2}[k]$ satisfying *both* of the following properties:
 - i. they must be contained in A ; and
 - ii. they must intersect *both* B and $\overline{\Theta_1}$.

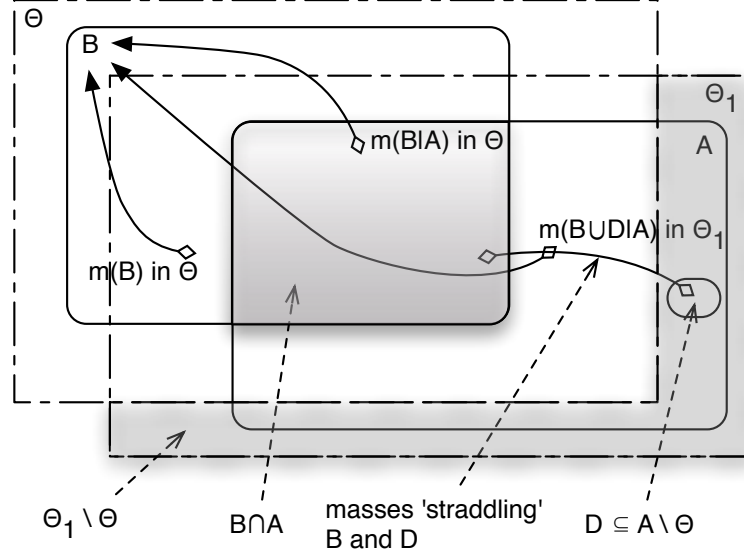


Figure 3.1: Updating the BoE $\mathcal{E}_{\Theta_1}[k] = \{\Theta_1, \mathfrak{F}_{\Theta_1}[k], m_{\Theta_1}(\cdot)[k]\}$ with the evidence of BoE $\mathcal{E}_{\Theta_2}[k] = \{\Theta_2, \mathfrak{F}_{\Theta_2}[k], m_{\Theta_2}(\cdot)[k]\}$ when $\Theta_1 \neq \Theta_2$ and $\Theta_1 \cap \Theta_2 \neq \emptyset$. The terms that contribute towards the update of $m_{\Theta_1}(B)[k]$ are shown.

These features can be extremely useful in some applications (e.g., soft/hard data fusion). In fact, we make use of this approach to define a new set of conditional notions that are applicable to non-identical FoDs (see Section 5.1 for details), based upon which a new soft hard fusion strategy is proposed.

3.4.6 Evidence Discounting

When an evidence source is not fully reliable⁸, a “discounting” operation can be performed on the associated mass function via [28]:

$$m_{\Theta}(B) = \begin{cases} d m_{\Theta}(B), & \text{for } B \subset \Theta; \\ 1 - d + d m_{\Theta}(B), & \text{for } B = \Theta, \end{cases} \quad (3.14)$$

⁸Refer to Chapter 7 for a detailed discussion on accounting for source imperfections, including the case of sources with unknown reliability.

where $d \in [0, 1]$ is referred to as the *discounting factor*. Often, the source reliability is used to discount evidence before performing fusion operations, for instance, evidence combinations.

3.4.7 Distance Measures for DST BoEs

A distance measure can be used to compare two given BoEs, similar to comparing two probability mass functions, for instance, via the *Kullback-Leibler (KL) divergence* [75]. Among many other options, a meaningful distance measure that takes into account the set-theoretic overlap of propositions is given by [76]:

Definition 10 (JGB Distance). *The distance between two BoEs \mathcal{E}_{Θ_i} s.t. $\Theta_i = \Theta$, for $i = 1, 2$, is given by*

$$\text{dist}(\mathcal{E}_{\Theta_1}, \mathcal{E}_{\Theta_2}) = \sqrt{\frac{1}{2} (\mathbf{m}_{\Theta_1} - \mathbf{m}_{\Theta_2})^T \mathcal{D}_{\Theta} (\mathbf{m}_{\Theta_1} - \mathbf{m}_{\Theta_2})}, \quad (3.15)$$

where $\mathbf{m}_{\Theta_i} = \{m_{\Theta_i}(\cdot)\}$, $i = 1, 2$, are $2^{\Theta} \times 1$ column vectors; and $\mathcal{D}_{\Theta} = \{d_{j\ell}\}$ is a $2^{\Theta} \times 2^{\Theta}$ matrix with $d_{j\ell} = |A_j \cap A_{\ell}| / |A_j \cup A_{\ell}|$, $A_j, A_{\ell} \in 2^{\Theta}$, $|\emptyset \cap \emptyset| / |\emptyset \cup \emptyset| \equiv 0$. \square

3.5 Chapter Summary

Different types of data uncertainties are generated depending on the how the data is originated and various imperfections of the source. Uncertainty handling formalisms provide techniques for modeling and working with these uncertain data types. For a given task, one needs to choose such formalisms that can properly capture and process the uncertainties typical of the task, yet satisfying the imposed constraints.

Part II

SOFT/HARD DATA FUSION FRAMEWORK

DST Conditional Notions

CONDITIONING, as in the case of probability theory, is the primary tool for DST evidence updating [56]. In fact, the core of the proposed soft/hard fusion framework is based on DST conditionals. Hence, it is essential that we fully understand the various conditional notions and their implications in order to be able to use them properly in various fusion operations.

We present a discussion on two of the most popular DST conditional notions, viz., DS conditionals [28] and FH conditionals [69], where we make the claim that the latter is the more appropriate choice for the task of soft/hard data fusion operations. However, due to the way FH conditionals are defined, identification of the conditional core is extremely difficult. This hampers the ability to properly understand the results of conditioning as well as the FH conditionals itself. The main result presented in this chapter redresses this shortcoming by providing a complete characterization of the conditional core generated by FH conditionals. In addition to providing insight into conditioning (and fusion operations based upon conditionals), one direct contribution of the proposed theorem is that it establishes the theoretical grounds for developing algorithms for efficient computation of FH conditionals. We also address the converse problem: “what events may have played a role in generating a given conditional core?”

This converse result can be of significant practical value in certain applications (e.g., for studying the sensitivity of the updated knowledge base with respect to the evidence received).

This chapter is organized as follows: Section 4.1 provides a detailed discussion on how DS and FH conditionals are defined, and the choice of an appropriate conditional notion for the task of soft/hard fusion; Section 4.2 contains our main result, its implications, the converse result along with numerical examples; Section 4.3 contains an application example illustrating the use of proposed theoretical results in a real life problem; Section 4.4 provides the concluding remarks. The proof of the main result together with several auxiliary results are relegated to Section A.1 in Appendix A.

4.1 Which Conditional Notion? And Why ?

Different conditional notions abound in DS theory literature. A rather comprehensive collection of DST conditional notions appear in a more recent article by Kimala and Yamada [77]. The question then is, what is the most appropriate conditional notion for the task at hand, viz., soft/hard data fusion?

The strategies that have been successfully implemented for hard data fusion are mainly Bayesian based. Moreover, there have been many attempts at relating DS theory to probability theory, including the original papers by Dempster [61, 68]. Given the abundant literature on probabilistic analysis of (hard) sensor networks, it is very beneficial to have a fusion strategy that merges well with existing and widely utilized strategies that work well with hard data. Based on this premise, we seek for conditional notions that offer a unique probabilistic interpretation leading to a more natural transition to the Bayesian conditional notions. Let us analyze, perhaps the two most popular, conditional notions in DST fusion literature.

4.1.1 DS Conditionals Versus FH Conditionals

The DS conditionals are developed on the following premise. Suppose we are given a mass function $m(\cdot)$ and the new evidence which states A has occurred. Then, if we model this new evidence via the mass function $m_A(\cdot)$, which has A as its only focal element, then the DS conditional with respect to A is defined as $m(\cdot | A) = (m \oplus m_A)(\cdot)$, where \oplus denotes the DCR. In fact, this is really the impetus behind the *definition* of the DS conditional; it is not necessarily an idea of a generalization of the Bayesian conditional [28]. As a result, DS conditionals are not appropriate for some applications, for instance, in soft/hard fusion where one may frequently encounter contradictory evidence, one specific case where the DCR can give counter-intuitive results. Thus, DS conditional notion only extend Bayesian conditioning in the sense that, when the focal elements are constituted of singletons only, it coincides with the Bayesian conditional. But then most, if not all, the conditional notions (including the FH conditional notion) possess this same property.

On the other hand, the FH conditional notion is that it can truly be considered a generalization of the Bayesian notion. This issue receives a comprehensive and rigorous treatment in [9, 69, 78]. Recall that, the inner and outer measures are respectively the best estimates one can make from below and above about the probability of a non-measurable event (for which a probability is not assigned). It turns out that the inner and outer measures induced by a probability function are indeed the belief and plausibility functions, respectively; and conversely, belief and plausibility functions are respectively the inner and outer measures induced by a probability function. It is this very precise relationship between inner and outer measures induced by a probability function and DS theoretic belief and plausibility functions that allows us to view DS theory as a generalization of, and being firmly rooted in, classical Bayesian probability [9, 78]. As demonstrated in the works of Fagin and Halpern [9, 69, 78], this is exactly what the FH conditional achieves. To our knowledge, the FH conditional

is the only DS conditional notion that can boast this property. In addition, FH conditionals avoid certain paradoxes associated with the DS conditional, e.g., the *sure thing principle* (see [2, 61, 79, 80], and in particular, [69] for a detailed discussion).

These are the reasons for opting to use the FH conditionals in our work. In what follows, unless otherwise mentioned, we restrict our attention to the FH conditionals only.

4.2 Characterization of the Conditional Core

In this section, we present the main result of this chapter. Let us proceed by introducing a special set construction and some preliminary notions, which are very useful for understanding the developments to follow.

4.2.1 Preliminaries

Definition 11 (Inner Sets and Outer Sets). *Consider a BoE $\mathcal{E}_\Theta \equiv \{\Theta, \mathfrak{F}_\Theta, m_\Theta(\cdot)\}$ and a conditioning event $A \subseteq \Theta$ s.t. $A \in \hat{\mathfrak{F}}_\Theta$. Then, the inner and outer sets of the conditioning event A are defined as*

$$\begin{aligned} in(A) &= \{B \subseteq A \mid B \in \mathfrak{F}_\Theta\}; \quad \text{and} \\ out(A) &= \{B \subseteq A \mid B \cup C \in \mathfrak{F}_\Theta, \emptyset \neq B, \emptyset \neq C \subseteq \overline{A}\}, \end{aligned}$$

respectively. The collections containing arbitrary unions of elements of $in(A)$ and $out(A)$ are given by

$$\begin{aligned} IN(A) &= \left\{ B \subseteq A \mid B = \bigcup_{i \in \mathcal{I}} C_i, C_i \in in(A) \right\}; \quad \text{and} \\ OUT(A) &= \left\{ B \subseteq A \mid B = \bigcup_{j \in \mathcal{J}} C_j, C_j \in out(A) \right\}, \end{aligned}$$

respectively; where, \mathcal{I} and \mathcal{J} are index sets that span the elements of $\text{in}(A)$ and $\text{out}(A)$, respectively. \square

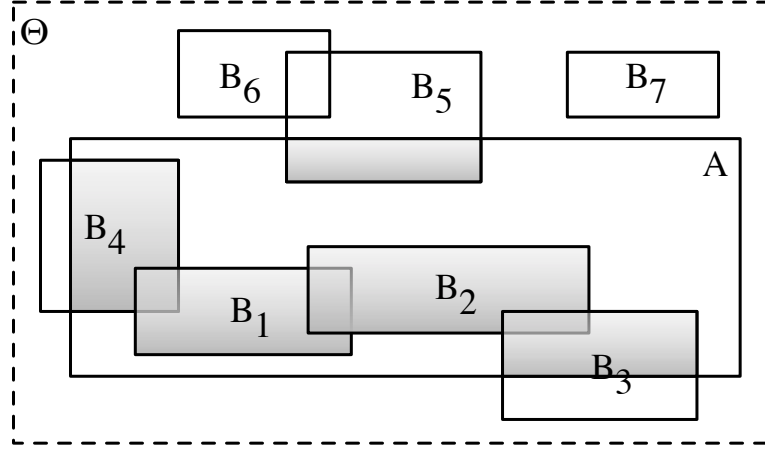


Figure 4.1: A set theoretic interpretation of $\text{in}(A)$ and $\text{out}(A)$ in Definition 11.

The set $\text{in}(A)$ contains all the focal elements that are *contained in* A ; $\text{out}(A)$ contains all the focal elements that intersect but are *not contained in* A . The elements of $\text{IN}(A)$ and $\text{OUT}(A)$ are arbitrary unions of elements of $\text{in}(A)$ and $\text{out}(A)$, respectively (See Fig. 4.1). As we will see later, the elements of $\text{in}(A)$ in union with elements of $\text{OUT}(A)$ form elements that are of special significance, i.e., the sets $B \subseteq A$, such that $B = X \cup Y$, for some $X \in \text{in}(A)$, $Y \in \text{OUT}(A)$. For B to be expressed in this manner, we must of course have $X \subseteq B$ and $Y \subseteq B$.

Definition 12 (Cumulative Mass). For a BoE $\mathcal{E}_\Theta \equiv \{\Theta, \mathfrak{F}_\Theta, m_\Theta(\cdot)\}$ and two arbitrary subsets $A \subseteq \Theta$ and $B \subseteq \Theta$, the sum

$$\mathcal{S}(A; B) = \sum_{\substack{\emptyset \neq X \subseteq A; \\ \emptyset \neq Y \subseteq B}} m_\Theta(X \cup Y),$$

denotes the cumulative mass of propositions that “straddle” A and B . \square

Remark: Clearly, for any $C \subseteq B \subseteq \Theta$, we have $\mathcal{S}(A; C) \leq \mathcal{S}(A; B)$.

With the definition of $\mathcal{S}(A; B)$ in place, we can now express the FH conditional in Theorem 2 as

Claim 3. *The conditional belief of any arbitrary proposition $B \subseteq \Theta$ is given by*

$$\text{Bl}_\Theta(B|A) = \frac{\text{Bl}_\Theta(A \cap B)}{\text{Pl}_\Theta(A) - \mathcal{S}(\bar{A}; A \cap B)}, \quad (4.1)$$

where A is the conditioning event s.t. $A \in \hat{\mathfrak{F}}_\Theta$. □

Proof. This is immediate from the relationship $\text{Pl}_\Theta(A) - \mathcal{S}(\bar{A}; A \cap B) = \text{Bl}_\Theta(A \cap B) + \text{Pl}_\Theta(A \cap \bar{B})$. ■

4.2.2 Conditional Core Theorem

We are now in a position to state the main result of this chapter, which is a theorem that establishes the necessary and sufficient conditions for a given proposition to belong to the conditional core.

Theorem 4 (Conditional Core Theorem (CCT)). *Let $\mathcal{E}_\Theta = \{\Theta, \mathfrak{F}_\Theta, m_\Theta(\cdot)\}$ be any arbitrary BoE. Then, the conditional mass function $m_\Theta(\cdot|A)$ satisfies*

$$m_\Theta(B|A) > 0 \iff B \in \text{in}(A) \quad \text{or} \quad B \in \text{in}(A) \cup \text{OUT}(A); \quad (4.2)$$

where $A \subseteq \Theta$ is any arbitrary conditioning event s.t. $A \in \hat{\mathfrak{F}}_\Theta$. □

Proof. See Section A.1 in Appendix A. ■

The proof of the CCT is somewhat laborious and hence we relegate it to the Appendix for the sake of clarity of the presentation. While at a first glance this proof may appear somewhat cumbersome, it is not too difficult to follow and it helps in understanding the form of the conditional focal elements and how they are generated.

The mathematical rigor and all the machinery involved are required to capture the subtleties that can arise depending on the “structure” of the core.

Let us illustrate the application of the CCT via a simple example.

Example 1. [52] *Consider a situation assessment scenario, where objects crossing the perimeter of a military facility are to be identified. Objects of concern are the following:*

$$F \equiv \text{Fighter}; M \equiv \text{Bomber}; T \equiv \text{Tank}; S \equiv \text{Soldier}; O \equiv \text{Other},$$

Each class of objects, except O , which cannot be further sub-classified, may further be sub-classified into either $f \equiv \text{friendly}$ or $e \equiv \text{enemy}$. Therefore, the total set of objects is

$$\Theta = \{\underbrace{F_e, F_f}_F, \underbrace{M_e, M_f}_M, \underbrace{T_e, T_f}_T, \underbrace{S_e, S_f}_S, O\}.$$

The BoE $\mathcal{E}_\Theta = \{\Theta, \mathfrak{F}_\Theta, m_\Theta(\cdot)\}$ represents the currently available evidence. Take

$$\mathfrak{F}_\Theta = \{M_e, M_f, S_f, (F_e, S_e), (M_e, T_e, O), \Theta\};$$

$$m_\Theta(B) = \{0.1, 0.1, 0.1, 0.2, 0.2, 0.3\},$$

for $B \in \mathfrak{F}_\Theta$ in the same order as in \mathfrak{F}_Θ .

Suppose the BoE \mathcal{E} needs to be updated to reflect new ground intelligence (which informs of perhaps an aerial object) by conditioning with respect to the conditioning

event $A = (F, M, O)$. First, we identify

$$\begin{aligned} in(A) &= \{M_e, M_f\}; \\ out(A) &= \{F_e, (M_e, O), (F, M, O)\}; \\ IN(A) &= \{M_e, M_f, M\}; \\ OUT(A) &= \{F_e, (M_e, O), (F_e, M_e, O), (F, M, O)\}. \end{aligned}$$

The only propositions that can be expressed as $X \cup Y$, $X \in in(A)$, $Y \in OUT(A)$, are

$$\mathcal{B} = \{(F_e, M_e), (F_e, M_f), (M_e, O), (F_e, M_e, O), (M, O), (F_e, M, O), (F, M, O)\}.$$

So, according to the CCT, $\mathfrak{F}_{\Theta|A}$ is the collection $\{X \subseteq \Theta \mid X \in in(A) \text{ or } X \in \mathcal{B}\}$ —the propositions contained in sets $in(A)$ and \mathcal{B} . Table 4.1 confirms this result.

B	$m_{\Theta}(B A)$	B	$m_{\Theta}(B A)$
M_f	0.11110	(F_e, M_e, O)	0.02540
M_e	0.11110	(M, O)	0.03175
(F_e, M_e)	0.03175	(F_e, M, O)	0.02540
(F_e, M_f)	0.03175	A	0.60000
(M_e, O)	0.03175	<i>all others</i>	0.00000

Table 4.1: Conditional masses generated by explicitly computing the conditional mass function with respect to the conditioning event $A = (F_f, F_e, M_f, M_e, O)$ in Example 1. Only the conditional focal elements, i.e., propositions with positive mass, are shown.

The following observations are noteworthy and make intuitive sense:

- Upon conditioning,
 - propositions supporting non-aerial objects do not remain in the conditional core;
 - propositions supporting only aerial objects remain in the conditional core.

- *The newly generated conditional focal elements can also be given intuitive interpretations.*
 - *While (F_e, S_e) goes to zero upon conditioning, the evidence existing toward F_e moves into the two propositions (F_e, M_e) and (F_e, M_f) .*
 - *The two propositions (F_e, S_e) and (M_e, T_e, O) goto zero upon conditioning. The evidence existing toward F_e in (F_e, S_e) does not move into F_e alone since there is no evidence toward a ‘singleton’ proposition. However, the support existing toward F_e and (M_e, O) moves to (M, F_e, O) .*

Thus, using the CCT, one can easily identify the propositions generated by conditioning without any numerical computations. As we have observed, these conditional focal elements can also be given an intuitive interpretation with respect to the available evidence and the conditioning event. ■

4.2.3 Implications of the CCT

Keeping in mind the observations made in Example 1, we now point out the several important implications of the CCT. Consider any arbitrary proposition $B \subseteq \Theta$ that is being conditioned with respect to an event $A \in \hat{\mathfrak{F}}_\Theta$.

- Suppose B is not contained in A : then B cannot belong to the conditional core.
- Suppose B is contained in A . Then,
 - if B belongs to the core, B belongs to the conditional core too;
 - if B does not belong to the core, B belongs to the conditional core iff it can be expressed as the union of a focal element contained in A and the

intersection of A with some arbitrary set of focal elements each of which straddles A and \overline{A} .

- If there are no focal elements that straddle A and \overline{A} , then the conditional core is equivalent to the core.
- Propositions with zero belief do not belong to the conditional core.

4.2.4 Converse of the CCT

Having received new evidence, suppose the BoE $\mathcal{E}_\Theta = \{\Theta, \mathfrak{F}_\Theta, m_\Theta(\cdot)\}$ gets updated to $\mathcal{E}'_\Theta = \{\Theta, \mathfrak{F}'_\Theta, m'_\Theta(\cdot)\}$. Knowledge of the conditioning proposition(s) that could have generated this updated knowledge can give us valuable clues as to the sensitivity of the knowledge base with respect to the evidence received. The following result—which can be considered a converse to the CCT—addresses this issue by “*bounding*” the sets of potential conditioning events, most importantly, with no recourse to numerical computations.

Lemma 5 (Conditional Core Generator (CCG) Bounds). *Let the BoE \mathcal{E}_Θ be updated to \mathcal{E}'_Θ after being conditioned with respect to $A \in \hat{\mathfrak{F}}_\Theta$, i.e., $\mathfrak{F}'_\Theta = \mathfrak{F}_{\Theta|A}$ and $m'_\Theta(B) = m_\Theta(B|A)$, $\forall B \subseteq \Theta$. Then, \mathfrak{F}'_Θ could have been generated from any conditioning proposition A that is bounded as $A_* \subseteq A \subseteq A^*$, where*

- (i) $\bigcup_{B \in \mathfrak{F}'_\Theta} B \subseteq A_*$,
- (ii) $\bigcup_{B \in \mathfrak{F}_\Theta} (B \setminus \mathcal{C}) \subseteq \overline{A^*}$, where $\mathcal{C} = \bigcup_{C \in \mathfrak{F}'_\Theta} C$.

Proof. directly follows from the CCT. ■

Let us consider the Example 1 again.

Example 2. For \mathfrak{F}_Θ and $\mathfrak{F}'_\Theta = \mathfrak{F}_{\Theta|A}$ in Example 1, we have $(F, M, O) \subseteq A_*$ and $(T, S) \subseteq \overline{A^*}$, thus making $A^* \subseteq (F, M, O)$. Hence, in this particular case, we must have the conditioning proposition $A = (F, M, O)$.

On the other hand, suppose we have

$$\mathfrak{F}_\Theta = \{M_f, M_e, S_f, (F_e, S_e), (M_e, T_e, O)\};$$

$$\mathfrak{F}'_\Theta = \{M_f, M_e, (F_e, M_f), (F_e, M_e), (M_e, O), (M, O), (F_e, M, O), (F_e, M_e, O)\}.$$

Then, $(F_e, M, O) \subseteq A_*$ and $(T_e, S) \subseteq \overline{A^*}$, thus making $A^* \subseteq (F, M, T_F, O)$. Hence, any conditioning proposition A satisfying $(F_e, M, O) \subseteq A \subseteq (F, M, T_F, O)$ may have generated this \mathfrak{F}'_Θ from \mathfrak{F}_Θ . ■

4.3 Application Example

Let us study the Target Identification (Target ID) case study that appears in [81], where several sensor reports are to be combined to determine the target ID in an air surveillance context. We use the same example, but with an intermediate evidence conditioning step to illustrate the application of the CCT and its converse.

4.3.1 Model

Suppose the *target allegiance* and *target class* FoDs are $\Theta_{\mathbb{A}} = \{f, n, s, h\}$ ($f = \text{friend}$, $n = \text{nueutral}$, $s = \text{suspect}$, $h = \text{hostile}$) and $\Theta_{\mathbb{B}'} = \{B_1, B_2, B_3, B_4\}$ ($B_1 = \text{Commercial Planes}$, $B_2 = \text{Fighter Planes}$, $B_3 = \text{Bombers}$, $B_4 = \text{Military Transport Planes}$), respectively. Each basic target class consists of several platform types: $B_1 = \{b_{11}, b_{12}\}$, $B_2 = \{b_{21}, b_{22}, b_{23}, b_{24}, b_{25}\}$, $B_3 = \{b_{31}, b_{32}, b_{33}\}$, and $B_4 = \{b_{41}, b_{42}\}$ (e.g., b_{11} could be Airbus-320). Hence, the FoD $\Theta_{\mathbb{B}} = B_1 \cup B_2 \cup B_3 \cup B_4$ is a refinement of $\Theta_{\mathbb{B}'}$. Target ID belongs to the FoD $\Theta = \Theta_{\mathbb{A}} \times \Theta_{\mathbb{B}}$.

Report ID	Corresponding Evidence BoE		
Report ⁽¹⁾ :	$m_{\Theta_{\mathbb{B}'}}^{(1)}(B_2) = 0.7;$	$m_{\Theta_{\mathbb{B}'}}^{(1)}(B_2, B_3) = 0.2;$	$m_{\Theta_{\mathbb{B}'}}^{(1)}(\mathbb{B}') = 0.1$
Report ⁽²⁾ :	$m_{\Theta_{\mathbb{B}}}^{(2)}(b_{22}) = 0.6;$	$m_{\Theta_{\mathbb{B}}}^{(2)}(b_{23}) = 0.3;$	$m_{\Theta_{\mathbb{B}}}^{(2)}(\mathbb{B}) = 0.1$
Report ⁽³⁾ :	$m_{\Theta_{\mathbb{A}}}^{(3)}(s) = 0.6;$	$m_{\Theta_{\mathbb{A}}}^{(3)}(\mathbb{A}) = 0.4$	
Report ⁽⁴⁾ :	$m_{\Theta_{\mathbb{A}}}^{(4)}(f) = 0.9;$	$m_{\Theta_{\mathbb{A}}}^{(4)}(\mathbb{A}) = 0.1$	

Table 4.2: Available Sensor Reports

4.3.2 Data Fusion and Decision-Making

The four sensor reports that are available are shown in Table 4.2. Fusion center combines all the sensor reports and generates a BoE for decision-making. Suppose the pairs $\{\mathbf{Report}^{(1)}, \mathbf{Report}^{(2)}\}$ and $\{\mathbf{Report}^{(3)}, \mathbf{Report}^{(4)}\}$ are fused separately.⁹ The resulting BoEs are vacuously extended and combined¹⁰ to obtain the fused mass $m_{\Theta}(\cdot)$ as

$$m_{\Theta}(\cdot) = \left(m_{\Theta_{\mathbb{B}'}}^{(1)}(\cdot) \oplus m_{\Theta_{\mathbb{B}}}^{(2)}(\cdot) \right)^{\mathbb{B} \uparrow \mathbb{A} \times \mathbb{B}} \oplus \left(m_{\Theta_{\mathbb{A}}}^{(3)}(\cdot) \oplus m_{\Theta_{\mathbb{A}}}^{(4)}(\cdot) \right)^{\mathbb{A} \uparrow \mathbb{A} \times \mathbb{B}}.$$

The target ID decision can be made by application of the pignistic transformation to $m_{\Theta}(\cdot)$ [33, 81]. Henceforth, when no confusion can occur, we avoid using the “comma” within non-singleton propositions (e.g., $(n, s, h) \times (B_1, B_2, B_3)$ is denoted as $nsh \times B_1 B_2 B_3$). Table 4.3 shows the fused results based on the sensor reports above and they appear to favor a target ID of $f \times b_{22}$ (i.e., a friendly b_{22}) [81].

4.3.3 Evidence Conditioning

Suppose high confidence intelligence confirms that neither **bombers** nor **friendly** or **commercial planes** are being deployed in the vicinity. The KB (i.e., $m_{\Theta}(\cdot)$) thus

⁹Sensor reports can be combined in any order using the DCR.

¹⁰See [81] for a detailed discussion.

B	$m_{\Theta}(B)$	B	$m_{\Theta}(B)$	B	$m_{\Theta}(B)$
$s \times b_{22}$	0.07824	$f \times b_{22}$	0.46956	$\Theta_{\mathbb{A}} \times b_{22}$	0.05220
$s \times b_{23}$	0.03912	$f \times b_{23}$	0.23478	$\Theta_{\mathbb{A}} \times b_{23}$	0.02610
$s \times B_2$	0.00913	$f \times B_2$	0.05478	$\Theta_{\mathbb{A}} \times B_2$	0.00609
$s \times B_2B_3$	0.00261	$f \times B_2B_3$	0.01565	$\Theta_{\mathbb{A}} \times B_2B_3$	0.00174
$s \times \Theta_{\mathbb{B}}$	0.00130	$f \times \Theta_{\mathbb{B}}$	0.00783	$\Theta_{\mathbb{A}} \times \Theta_{\mathbb{B}}$	0.00087

Table 4.3: Fused BoE Generated From the Sensor Reports

can be conditioned with respect to the event $A = nsh \times B_2B_4$. The associated computation can be efficiently carried out by using the CCT to first identify the conditional focal elements. Recall that the set $\text{in}(A)$ captures the focal elements that are contained in A ; $\text{out}(A)$ captures the focal elements that intersect but are not contained in A . So we get

$$\begin{aligned} \text{in}(A) &= \{s \times b_{22}, s \times b_{23}, s \times B_2\}; \\ \text{out}(A) &= \{s \times B_2, s \times B_2B_4, nsh \times b_{22}, nsh \times b_{23}, nsh \times B_2, nsh \times B_2B_4\}. \end{aligned}$$

Table 4.4 shows *all* the elements that are either elements of $\text{in}(A)$ or $\text{in}(A) \cup \text{OUT}(A)$. The CCT states that these are the *only* elements that belong to the conditional core $\mathfrak{F}_{\Theta|A}$. A direct computation confirms this conclusion; these conditional mass values are also indicated in Table 4.4. Similar to those in Example 1, notice the following on conditional propositions.

- Upon conditioning, no proposition supporting bombers (B_3), friendly (f) planes or commercial (B_1) planes are retained in the conditional core.
- The newly generated conditional focal elements can also be given an intuitive interpretation. Upon conditioning, the support for $s \times \Theta_{\mathbb{B}}$ vanishes and moves toward non-commercial non-bomber objects represented in $s \times B_2B_4$. The support

for $\Theta_{\mathbb{A}} \times b_{22}$ also vanishes. The support that had existed towards non-friendly planes b_{22} moves toward $nsh \times b_{22}$, but not toward focal elements such as $n \times b_{22}$ because there is no evidence supporting such “singleton” propositions.

B	$\text{Bl}(B A)$	$m(B A)$	B	$\text{Bl}(B A)$	$m(B A)$
$s \times b_{22}$	0.35989	0.35989	$s \times b_{23}$	0.17994	0.17994
$s \times B_2$	0.58890	0.04907	$s \times B_2B_4$	0.59249	0.00359
$nsh \times b_{22}$	0.47361	0.11372	$nsh \times b_{23}$	0.20450	0.02456
$(s \times b_{22}) \cup (nsh \times b_{23})$	0.61349	0.04910	$(s \times b_{23}) \cup (nsh \times b_{22})$	0.71041	0.05686
$nsh \times b_{22}b_{23}$	0.84371	0.05964	$(s \times B_2) \cup (nsh \times b_{22})$	0.77797	0.01849
$(s \times B_2) \cup (nsh \times b_{23})$	0.67036	0.00780	$(s \times B_2) \cup (nsh \times b_{22}b_{23})$	0.92673	0.00766
$nsh \times B_2$	0.98313	0.05640	$(s \times B_2B_4) \cup (nsh \times b_{22})$	0.78424	0.00268
$(s \times B_2B_4) \cup (nsh \times b_{23})$	0.67501	0.00106	$(s \times B_2B_4) \cup (nsh \times b_{22}b_{23})$	0.93565	0.00159
$(s \times B_2B_4) \cup (nsh \times B_2)$	0.99317	0.00112	$A = nsh \times B_2B_4$	1.00000	0.00683

Table 4.4: Conditional Core $\mathfrak{F}_{\Theta|A}$ Corresponding to $A = nsh \times B_2B_4$

Let us use the converse of the CCT on the conditioned BoE to identify potential conditioning events that may have generated the new (conditioned) BoE.

We have $A_* \supseteq nsh \times B_2B_3B_4$ and $\overline{A^*} \supseteq (nsh \times B_1, f \times \Theta_{\mathbb{B}})$ so that $A^* \subseteq nsh \times B_2B_3B_4$. Thus, in this particular case A can be uniquely determined¹¹ as $A = nsh \times B_2B_3B_4$.

On the other hand, suppose $m_{\Theta}(\Theta) = 0$ (and the other masses normalized) in the original mass assignment in Table 4.3. Then, $A_* \supseteq (nsh \times B_2B_3, s \times B_4)$ and $\overline{A^*} \supseteq (f \times \Theta_{\mathbb{B}}, s \times B_1)$, so that $A^* \subseteq (nsh \times B_2B_3B_4, nh \times B_1)$. Thus, the conditioning event A is bounded as $(nsh \times B_2B_3, s \times B_4) \subseteq A \subseteq (nsh \times B_2B_3B_4, nh \times B_1)$. In this

¹¹It is important to realize that for a given revised (conditioned) BoE, there can be many possible conditioning events that may have generated the revised BoE from the originally cast BoE. Hence, in general, the conditioning event cannot be determined uniquely.

case, we can only bound the potential conditioning event by two sets—a set that is *contained in* and another set that *contains*, the conditioning event.

4.4 Chapter Summary

There is no one conditional notion that is applicable to all the applications. We have identified the FH conditionals to be a more suitable candidate for the task of soft/hard fusion. The main result presented in this chapter is the CCT which provides a complete characterization of the conditional focal elements with no recourse to numerical computations. The CCT helps to better understand the conditional notions and fusion operations, such as evidence updating schemes, that are based upon them. In addition, the CCT lays the theoretical foundation for the development of computational schemes for efficient calculation of FH conditionals. The CCG bounds provide a way to bound the sets of conditioning events that may have caused a belief change, again without numerical computations. These results can be very helpful in understanding and interpreting dynamic changes in knowledge bases.

Conditional Approach to Data Fusion

A new DST fusion strategy that forms the core of the developed analytical framework is presented. It addresses the challenges \mathbf{C}_1 - \mathbf{C}_4 associated with soft/hard data fusion (see Chapter 1 for details). While satisfying a majority of the algebraic and fusion properties common to other widely used DST combination rules, this new strategy also possesses several other intuitively appealing features which makes it ideal for many soft/hard fusion applications. For instance, it is robust and provides more reasonable results when confronted with contradictory evidence, a feature that is lacking in perhaps the most widely used DST fusion rule—the DCR. We propose several fusion techniques and the associated parameter selection strategies that are applicable to a multitude of fusion tasks. These proposed techniques are further explained via a running example. We also provide a detailed example and two real-life application examples to illustrate various parameter selection and fusion strategies presented in this chapter.

This chapter is organized as follows: Section 5.1 presents new conditional operators that are applicable for sources having non-identical FoDs; Section 5.2 presents a new fusion strategy for soft/hard data based on the above conditional notions; analysis of various fusion characteristics of the presented method are also given; Section 5.3

presents several parameter selection strategies; Section 5.4 contains an illustrative example; Section 5.5 presents two real life applications exhibiting some of the key concepts and ideas introduced in the chapter; and finally, Section 5.6 contains the chapter summary. The proofs of the results that do not appear within the text are provided in Section A.2 of Appendix A.

5.1 Ported Conditional Notions

The conditional approach proposed in [74] is only applicable to evidence updating. However, the features such as

- (i) not having to expand one's own FoD, and
- (ii) the ability to refine and focus on relevant information via conditioning,

are indeed attractive properties in any fusion scenario, especially in the case of soft/hard applications. In the conditional approach to belief updating in (3.13) (see Section 3.4.5.3), the conditional belief is computed in \mathcal{E}_{Θ_2} (i.e., using the evidence in \mathcal{E}_{Θ_2}) for updating the belief of a proposition B in Θ_1 . One can view this as a belief computation in a BoE being “ported” to another BoE, where the two FoDs need not be identical. With this observation in mind, we can identify new conditional operators that are applicable to both identical and non-identical FoDs. This also allows for a unified representation of our fusion strategy. Let us proceed as follows.

Consider two BoEs $\mathcal{E}_{\Theta_i} = \{\Theta_i, \mathfrak{F}_{\Theta_i}, m_{\Theta_i}(\cdot)\}$, $i = 1, 2$ s.t. $\Theta_1 \cap \Theta_2 \neq \emptyset$ (i.e., Θ_1 and Θ_2 have at least one element in common, since, otherwise the fusion does not make sense). Also, for any conditioning event $A \in \hat{\mathfrak{F}}_{\Theta_2}$ s.t. $\text{Pl}_{\Theta_2}(\Theta_1|A) > 0$, define a normalizing constant $K_{\Theta_2}(\Theta_1|A) = \text{Bl}_{\Theta_2}(\Theta_1|A) + \text{Pl}_{\Theta_2}(\Theta_1|A)$.

5.1.1 Ported Conditional Belief Function

Definition 13 (Ported Conditional Belief). *The ported conditional belief of a proposition $B \subseteq \Theta_1$ computed in \mathcal{E}_{Θ_2} is given by*

$$\text{Bl}_{\Theta_1}^{(\Theta_2)}(B|A) = \frac{1}{K_{\Theta_2}(\Theta_1|A)} \left(\text{Bl}_{\Theta_2}(B|A) + \text{Bl}_{\Theta_2}(B \cup \Theta_{2 \setminus 1}|A) - \text{Bl}_{\Theta_2}(\Theta_{2 \setminus 1}|A) \right), \quad (5.1)$$

for any $A \in \hat{\mathfrak{F}}_{\Theta_2}$ s.t. $\text{Pl}_{\Theta_2}(\Theta_1|A) > 0$. □

Notice the following.

- Clearly, $\text{Bl}_{\Theta_1}^{(\Theta_2)}(B|A) = \text{Bl}_{\Theta_2}(B|A)$, when $\Theta_1 = \Theta_2$; hence, ported conditional belief function reduces to a regular belief function, when FoDs are identical.
- The ported conditional belief is not defined when $\text{Pl}_{\Theta_2}(\Theta_1|A) = 0$. In fact, if even Θ_1 is not plausible in \mathcal{E}_{Θ_2} when A is given, then it does not make any sense to compute conditional beliefs (or plausibility or BPA for that matter) in \mathcal{E}_{Θ_2} for any subset B of Θ_1 .

Claim 6. *The ported conditional belief $\text{Bl}_{\Theta_1}^{(\Theta_2)}(\cdot|A) : 2^{\Theta_1} \mapsto [0, 1]$ is a valid belief function on Θ_1 , whenever $A \in \hat{\mathfrak{F}}_{\Theta_2}$ and $\text{Pl}_{\Theta_2}(\Theta_1|A) > 0$. □*

Proof. See Section A.2.1 of Appendix A. ■

Example 3. *Consider the BoEs \mathcal{E}_{Θ_i} , $i = 1, 2$ where $\Theta_1 = \{a, b, c\}$, $\Theta_2 = \{b, c\}$ and $\{m_{\Theta_1}(a), m_{\Theta_1}(ab), m_{\Theta_1}(bc)\} = \{0.4, 0.4, 0.2\}$ and $\{m_{\Theta_2}(b), m_{\Theta_2}(bc)\} = \{0.8, 0.2\}$. Say, we would like to compute ported belief conditionals $\text{Bl}_{\Theta}^{(\Theta_i)}(\cdot|\cdot)$, $i = 1, 2$, for $\Theta = \{b, c\}$.*

Take the computation of $\text{Bl}_{\Theta}^{(\Theta_1)}(\cdot|ab)$ (Note that $\text{Pl}_{\Theta_1}(bc|ab) > 0$). First, compute the belief functions $\text{Bl}_{\Theta_1}(\cdot|ab)$ and obtain the normalization constant $K_1(\Theta|ab)$ as

B	$\text{Bl}_{\Theta_1}(B ab)$	$\text{Bl}_{\Theta}^{(\Theta_1)}(B ab)$	$\text{Bl}_{\Theta_2}(B bc)$	$\text{Bl}_{\Theta}^{(\Theta_2)}(B bc)$
\emptyset	-	-	-	-
a	0.1	-	-	-
b	-	1.0	0.8	0.8
c	-	-	-	-
ab	1.0	-	-	-
ac	0.1	-	-	-
bc	-	1.0	1.0	1.0
abc	1.0	-	-	-

Table 5.1: Computation Example of Ported Conditional Beliefs

$K_i(\Theta|ab) = \text{Bl}_{\Theta_1}(\Theta|ab) + \text{Pl}_{\Theta_1}(\Theta|ab) = 1 + \text{Bl}_{\Theta_1}(\Theta|ab) - \text{Bl}_{\Theta_1}(\Theta_1 \setminus \Theta|ab)$. Then, use (5.1) to obtain $\text{Bl}_{\Theta}^{(\Theta_1)}(\cdot|ab)$ (see Table 5.1). Notice that $\text{Bl}_{\Theta}^{(\Theta_2)}(\cdot|bc)$ is identical to $\text{Bl}_{\Theta_2}(\cdot|bc)$, since $\Theta_2 = \Theta$. ■

5.1.2 Ported Conditional Plausibility Function

Since the ported conditional belief in Definition 13 turns out to be a valid DST belief function, we can now use the standard definition to derive *ported conditional plausibility* as:

Definition 14 (Ported Conditional Plausibility). *The ported conditional plausibility of a proposition $B \subseteq \Theta_1$ is*

$$\text{Pl}_{\Theta_1}^{(\Theta_2)}(B|A) = 1 - \text{Bl}_{\Theta_1}^{(\Theta_2)}(\Theta_1 \setminus B|A), \quad (5.2)$$

for any $A \in \hat{\mathfrak{F}}_{\Theta_2}$ s.t. $\text{Pl}_{\Theta_2}(\Theta_1|A) > 0$. □

Claim 7. *The ported conditional plausibility function $\text{Pl}_{\Theta_1}^{(\Theta_2)}(\cdot|A) : 2^{\Theta_1} \mapsto [0, 1]$ defined in Definition 14 is given by*

$$\text{Pl}_{\Theta_1}^{(\Theta_2)}(B|A) = \frac{1}{K_{\Theta_2}(\Theta_1|A)} \left(\text{Pl}_{\Theta_2}(B|A) + \text{Pl}_{\Theta_2}(B \cup \Theta_{2 \setminus 1}|A) - \text{Pl}_{\Theta_2}(\Theta_{2 \setminus 1}|A) \right), \quad (5.3)$$

for any $A \in \hat{\mathfrak{F}}_{\Theta_2}$ s.t. $\text{Pl}_{\Theta_2}(\Theta_1|A) > 0$. □

Proof. by direct substitution of (5.2) into (5.1). ■

Notice that $\text{Pl}_{\Theta_1}^{(\Theta_2)}(B|A)$ also reduces to $\text{Pl}_{\Theta_2}(B|A)$, when $\Theta_1 = \Theta_2$.

5.1.3 Ported Conditional Mass Function

Similarly, we can define the *ported conditional mass* function using the standard belief to mass transformation as:

Definition 15 (Ported Conditional Mass). *The ported conditional mass of a proposition $B \subseteq \Theta_1$ is given by*

$$m_{\Theta_1}^{(\Theta_2)}(B|A) = \sum_{C \subseteq B} (-1)^{|B-C|} \text{Bl}_{\Theta_1}^{(\Theta_2)}(C|A), \quad (5.4)$$

for any $A \in \hat{\mathfrak{F}}_{\Theta_2}$ s.t. $\text{Pl}_{\Theta_2}(\Theta_1|A) > 0$. □

Claim 8. *The ported conditional mass function $m_{\Theta_1}^{(\Theta_2)}(\cdot|A) : 2^{\Theta_1} \mapsto [0, 1]$ in Definition 15 can be expressed as*

$$m_{\Theta_1}^{(\Theta_2)}(B|A) = \frac{m_{\Theta_2}(B|A) + \sum_{D \subseteq A \setminus \Theta_1} m_{\Theta_2}(B \cup D|A)}{K_{\Theta_2}(\Theta_1|A)}, \quad \text{for } \emptyset \neq B \subseteq \Theta_1;$$

with $m_{\Theta_1}^{(\Theta_2)}(\emptyset|A) = 0$, for any $A \in \hat{\mathfrak{F}}_{\Theta_2}$ s.t. $\text{Pl}_{\Theta_2}(\Theta_1|A) > 0$. □

Proof. See Section A.2.2 of Appendix A. ■

Note that, similar to the above cases, $m_{\Theta_1}^{(\Theta_2)}(B|A)$ also reduces to $m_{\Theta_2}(B|A)$, when $\Theta_1 = \Theta_2$. As shown in the next section, it is this very property of ported conditionals that allows us a unified representation of the fusion strategy in both identical and non-identical FoD cases. Moreover, due to the use of ported conditionals, as it turns out, the proposed fusion equation has the same functional form in all belief, plausibility and mass functions.

5.2 Conditional Fusion Equation

With the ported conditionals in place, now we are in a position to present our new evidence combination strategy.

Definition 16 (Conditional Fusion Equation (CFE)). *The belief function associated with the CFE-generated fused BoE $\mathcal{E}_\Theta = \mathcal{E}_{\Theta_1} \bowtie \mathcal{E}_{\Theta_2}$ is*

$$\text{Bl}_\Theta(B) = \sum_{A_1 \in \mathfrak{F}_{\Theta_1}} \beta_1(A_1) \text{Bl}_\Theta^{(\Theta_1)}(B|A_1) + \sum_{A_2 \in \mathfrak{F}_{\Theta_2}} \beta_2(A_2) \text{Bl}_\Theta^{(\Theta_2)}(B|A_2), \quad (5.5)$$

whenever $\text{Pl}_{\Theta_i}(\Theta|\Theta_i) > 0$, for $i = 1, 2$. Here, $\beta_i(\cdot)$, $i = 1, 2$ are non-negative, real and

$$\text{satisfy } \sum_{A_1 \in \mathfrak{F}_{\Theta_1}} \beta_1(A_1) + \sum_{A_2 \in \mathfrak{F}_{\Theta_2}} \beta_2(A_2) = 1. \quad \square$$

The CFE above defines a whole family of general fusion equations parameterized via $\beta_i(\cdot)$, $i = 1, 2$.

Remarks:

- The sets $A \in \mathfrak{F}_{\Theta_i}$ for which $\beta_i(A) > 0$, identify the conditioning events. Incoming evidence is further “conditioned” (or “refined”) based on these events.
- The linear combination weights $\beta_i(A) > 0$ can be used to emphasize or de-emphasize the contribution from each conditioned proposition.
- The desired FoD Θ , can be chosen depending on the application requirements.
The only condition CFE imposed on Θ is $\Theta \cap \Theta_i \neq \emptyset$, for $i=1, 2$.

In essence, CFE defines the belief of a proposition $B \subseteq \Theta$ as the weighted mean of conditional belief of the proposition B taken over the sets of focal elements in BoEs to be fused.

Claim 9. *The belief function $\text{Bl}_\Theta(\cdot) : 2^\Theta \mapsto [0, 1]$ given in CFE Definition 16 is a valid belief function on the FoD Θ .* \square

Proof. Note that $\text{Bl}_\Theta^{(\Theta_i)}(\cdot|A_i)$ is valid belief function on Θ for all $A \in \mathfrak{F}_{\Theta_i}$, for $i = 1, 2$. The claim follows from the fact that an arbitrary convex combination of belief functions defined on one FoD remains a valid belief function on the same FoD. \blacksquare

Example 4. *Consider the Example 3 again. Let us compute the fused BoE $\mathcal{E}_\Theta = \mathcal{E}_{\Theta_1} \bowtie \mathcal{E}_{\Theta_2}$ with parameters $\beta_i(A) = \mathcal{K}_i m_{\Theta_i}(A)$, where \mathcal{K}_i are normalization constants (this particular parameter selection is referred to as rCFE; see Section 5.3.3.1 for details). Table 5.2 lists the relevant ported conditional beliefs.*

B	$\text{Bl}_\Theta^{(\Theta_1)}(B ab)$	$\text{Bl}_\Theta^{(\Theta_1)}(B bc)$	$\text{Bl}_\Theta^{(\Theta_2)}(B b)$	$\text{Bl}_\Theta^{(\Theta_2)}(B bc)$	$\text{Bl}_\Theta(B)$
\emptyset	0.0	0.0	0.0	0.0	0.00
b	1.0	0.0	1.0	0.8	0.85
c	0.0	0.0	0.0	0.0	0.00
bc	1.0	1.0	1.0	1.0	1.00
	$\beta_1(ab) = .250$	$\beta_1(bc) = .125$	$\beta_2(b) = .500$	$\beta_2(bc) = .125$	

Table 5.2: Ported Conditional Belief Values for $\text{Bl}_\Theta(\cdot)$ Computation in Example 4.

Now, use (5.5) to compute the belief values. \blacksquare

We may now extract the plausibility and mass forms corresponding to the belief equation in the CFE. As it turns out, all of these equations have the same functional form.

Claim 10. *The plausibility and mass functions associated with the CFE-generated fused BoE $\mathcal{E}_\Theta = \mathcal{E}_{\Theta_1} \bowtie \mathcal{E}_{\Theta_2}$ are, respectively,*

$$\text{Pl}_\Theta(B) = \sum_{A_1 \in \mathfrak{F}_{\Theta_1}} \beta_1(A_1) \text{Pl}_\Theta^{(\Theta_1)}(B|A_1) + \sum_{A_2 \in \mathfrak{F}_{\Theta_2}} \beta_2(A_2) \text{Pl}_\Theta^{(\Theta_2)}(B|A_2); \quad (5.6)$$

$$m_\Theta(B) = \sum_{A_1 \in \mathfrak{F}_{\Theta_1}} \beta_1(A_1) m_\Theta^{(\Theta_1)}(B|A_1) + \sum_{A_2 \in \mathfrak{F}_{\Theta_2}} \beta_2(A_2) m_\Theta^{(\Theta_2)}(B|A_2); \quad (5.7)$$

whenever $\text{Pl}_{\Theta_i}(\Theta|\Theta_i) > 0$, for $i = 1, 2$. Here, $\beta_i(\cdot)$, $i = 1, 2$ are non-negative, real and satisfy $\sum_{A_1 \in \mathfrak{F}_{\Theta_1}} \beta_1(A_1) + \sum_{A_2 \in \mathfrak{F}_{\Theta_2}} \beta_2(A_2) = 1$. \square

Proof. The plausibility equation in (5.6) can be readily obtained by direct substitution of identity $\text{Pl}_\Theta(B) = 1 - \text{Bl}_\Theta(\Theta \setminus B)$ into (5.5) in CFE Definition. \square

To get the mass equation, expand the right-hand side (RHS) of (5.5) in terms of masses which yields

$$\text{RHS} = \sum_{C \subseteq B} \left(\sum_{A_1 \in \mathfrak{F}_{\Theta_1}} \beta_1(A_1) m_\Theta^{(\Theta_1)}(C|A_1) + \sum_{A_2 \in \mathfrak{F}_{\Theta_2}} \beta_2(A_2) m_\Theta^{(\Theta_2)}(C|A_2) \right) \quad (5.8)$$

But, the left-hand side (LHS) of (5.5) yields

$$\text{LHS} = \sum_{C \subseteq B} m(C) \quad (5.9)$$

Equate the LHS and RHS to get

$$\sum_{C \subseteq B} \left\{ m_\Theta(B) - \left(\sum_{A_1 \in \mathfrak{F}_{\Theta_1}} \beta_1(A_1) m_\Theta^{(\Theta_1)}(C|A_1) + \sum_{A_2 \in \mathfrak{F}_{\Theta_2}} \beta_2(A_2) m_\Theta^{(\Theta_2)}(C|A_2) \right) \right\} = 0 \quad (5.10)$$

Since, $B \subseteq \Theta$ is arbitrary, we must have

$$m_\Theta(B) - \left(\sum_{A_1 \in \mathfrak{F}_{\Theta_1}} \beta_1(A_1) m_\Theta^{(\Theta_1)}(C|A_1) + \sum_{A_2 \in \mathfrak{F}_{\Theta_2}} \beta_2(A_2) m_\Theta^{(\Theta_2)}(C|A_2) \right) = 0, \quad (5.11)$$

for all $B \subseteq \Theta$. Hence, the claim. \blacksquare

5.2.1 Properties of the CFE

Let us study some the important properties of CFE in order to better understand how it behaves under various fusion conditions.

5.2.1.1 Algebraic Properties

Commutativity: This is trivial from the definition.

Associativity: Arbitrary fusion with the CFE does not preserve associativity, in general. However, CFE can be used for associativity preserving fusion by, either using the multiple-BoE form or selecting an appropriate set of parameters (see Sections 5.2.3.1 and 5.2.3.2, respectively).

Continuity: In the context of evidence fusion, the continuity principle implies that a small change in the BoEs being fused should cause only a small change in the fused result [3]. This is definitely true for the CFE because conditionals are continuous and the CFE itself is a convex combination of conditionals. So, the CFE-generated fused BoE will undergo only a small change provided that the sets \mathfrak{F}_{Θ_i} remain unchanged with small changes in \mathcal{E}_{Θ_i} , for $i = 1, 2$.

Idempotency: CFE is not idempotent, in general. However, it can be shown that, if \mathfrak{F}_{Θ_i} is s.t. every $A_i \in \mathfrak{F}_{\Theta_i}$ contains all the focal elements in \mathcal{E}_{Θ_i} , for $i = 1, 2$, then the CFE is idempotent. For instance, the trivial case results when $\beta_i(A_i) = 0, \forall A_i \neq \Theta_i$, for $i = 1, 2$.

5.2.1.2 Fusion of Vacuous BoEs

Fusion of two vacuous BoEs, in the case of non-identical FoDs, is an interesting, however not very intuitive situation. The following example illustrates this non-trivial nature of fusing two vacuous BoEs, when the two corresponding FoDs are non-identical.

Example 5. [52] Consider the fusion of \mathcal{E}_{Θ_1} and \mathcal{E}_{Θ_2} , where \mathcal{E}_{Θ_1} is vacuous, $\Theta_1 = \{ab\}$ and $\Theta_2 = \{bc \dots\}$. In the cases of, (a) $m_{\Theta_2}(bc) = 1.0$, (b) $m_{\Theta_2}(bcde) = 1.0$ and (c) $m_{\Theta_2}(\Theta_2) = 1.0$, when should we expect the fused BoE \mathcal{E}_{Θ} to be vacuous? \square

How the CFE handles such a situation becomes clear via:

Claim 11. Consider the CFE-generated fused BoE $\mathcal{E}_{\Theta} = \mathcal{E}_{\Theta_1} \bowtie \mathcal{E}_{\Theta_2}$, where \mathcal{E}_{Θ_i} , $i = 1, 2$, are both vacuous. Then,

(i) if $\Theta \subseteq \Theta_1 \cap \Theta_2$, \mathcal{E}_{Θ} is vacuous;

(ii) otherwise, $m_{\Theta}(\Theta_i \cap \Theta) = \beta_i(\Theta_i)$, $i = 1, 2$. \square

Proof. Since \mathcal{E}_{Θ_i} is vacuous, we have $\mathfrak{F}_{\Theta_i} = \{\Theta_i\}$; therefore, the parameters $\beta_i(B) = 0$, $\forall B \subset \Theta_i$, for $i = 1, 2$. For a non-trivial fusion, assume $\beta_i(\Theta_i) \neq 0$, $i = 1, 2$. Then the mass update equation reduces to

$$m_{\Theta}(B) = \beta_1(\Theta_1) m_{\Theta}^{(\Theta_1)}(B|\Theta_1) + \beta_2(\Theta_2) m_{\Theta}^{(\Theta_2)}(B|\Theta_2).$$

Use the CCT in Theorem 4, to obtain the result. \blacksquare

5.2.1.3 Fusion of Contradictory Evidence

Fusion of contradictory evidence is one of the most crucial challenges in soft/hard fusion. Some of the well established DST combination rules (e.g., the DCR) are known to yield counter-intuitive results in the presence of contradictory evidence.

Let us study the behavior of CFE in this situation. For illustrational purposes, consider the fusion $\mathcal{E}_{\Theta} = \mathcal{E}_{\Theta_1} \bowtie \mathcal{E}_{\Theta_2}$, where

$$\mathcal{E}_{\Theta_1} : \{m_{\Theta_1}(a), m_{\Theta_1}(\Theta_1)\} = \{\mu, 1 - \mu\};$$

$$\mathcal{E}_{\Theta_2} : \{m_{\Theta_2}(b), m_{\Theta_2}(\Theta_2)\} = \{\nu, 1 - \nu\}, \nu < \mu < 1,$$

with $\Theta \equiv \Theta_1 = \Theta_2 = \{a, b\}$. The CFE generates

$$\begin{aligned} m_{\Theta}(a) &= \beta_1(a) + \beta_1(\Theta) \mu, & m_{\Theta}(b) &= \beta_2(b) + \beta_2(\Theta) \nu, \\ m_{\Theta}(\Theta) &= \beta_1(\Theta) (1 - \mu) + \beta_2(\Theta) (1 - \nu), \end{aligned}$$

where $\beta_1(a) + \beta_1(\Theta) + \beta_2(b) + \beta_2(\Theta) = 1$. Clearly, the results are parameter dependent. A particularly interesting parameter configuration that we refer to as *rCFE* (see Section 5.3 for details) yields the ‘odds ratio’

$$\frac{m_{\Theta}(a)}{m_{\Theta}(b)} = \frac{\beta_1(\Theta)}{1 - \beta_1(\Theta)} \frac{\mu}{\nu} \rightarrow \frac{\beta_1(\Theta)}{1 - \beta_1(\Theta)}, \quad \text{as } \nu \rightarrow 1 \text{ with } \nu < \mu < 1.$$

In contrast, the odds ratio generated from most of the combination rules based on the conjunctive form, for instance DCR, tends to ∞ (irrespective of how close to one ν is). Clearly, the CFE behaves more reasonably in this scenario. Let us illustrate the behavior of several popular DST combination rules via two well-known examples. In applying the rCFE and cCFE examples, we use $\mathcal{K}_1 : \mathcal{K}_2 = 1 : 1$ (see Section 5.3 for details on parameter selection). We also use the notation $\theta_{ij} = (\theta_i, \theta_j)$ and $\theta_{123} = \{\theta_1, \theta_2, \theta_3\}$. The first example is due to Zadeh [2].

Example 6. [2] *Let us fuse \mathcal{E}_{Θ_1} and \mathcal{E}_{Θ_2} , where $\Theta_1 = \Theta_2 \equiv \Theta = \{\theta_1, \theta_2, \theta_3\}$ and $\{m_{\Theta_1}(\theta_1), m_{\Theta_1}(\theta_3)\} = \{0.9, 0.1\}$, $\{m_{\Theta_2}(\theta_2), m_{\Theta_2}(\theta_3)\} = \{0.9, 0.1\}$. The results generated by the CFE and other combination rules appear in Table 5.3.*

The rCFE simulates the scenario when one needs to aggregate the information, resulting in an output similar to the conjunctive rule and its derivatives. On the other hand, the cCFE simulates the scenario where one looks at evidence that both sources “agree” on. Hence, even though the evidence towards θ_3 is low, it is the only proposition both \mathcal{E}_{Θ_1} and \mathcal{E}_{Θ_2} can agree upon. The result generated in this case is similar to the DCR and DP2. □

$B \in \mathfrak{F}$	<i>Disj.</i>	<i>Conj.</i>	<i>DCR</i>	<i>YGR</i>	<i>DP1</i>	<i>DP2</i>	<i>PCR</i>	<i>RCR</i>	<i>rCFE</i>	<i>cCFE</i>
\emptyset	-	0.99	-	-	-	-	-	-	-	-
θ_1	-	-	-	-	-	-	0.49	-	0.45	-
θ_2	-	-	-	-	-	-	0.49	-	0.45	-
θ_3	0.01	0.01	1.00	0.01	0.01	0.50	0.02	0.01	0.10	1.00
θ_{12}	0.81	-	-	-	0.81	0.40	-	0.81	-	-
θ_{13}	0.09	-	-	-	0.09	0.05	-	0.09	-	-
θ_{23}	0.09	-	-	-	0.09	0.05	-	0.09	-	-
Θ	-	-	-	0.99	0.09	-	-	-	-	-
<i>Decision</i>	θ_{12}	\emptyset	θ_3	Θ	θ_{12}	θ_3	θ_{12}	θ_{12}	θ_{12}	θ_3

Table 5.3: Fusion of Contradictory Evidence: Example from [2]

The example due to Florea, et al [3] is:

Example 7. [3] Let us fuse \mathcal{E}_{Θ_1} and \mathcal{E}_{Θ_2} , where $\Theta_1 = \Theta_2 \equiv \Theta = \{\theta_1, \dots, \theta_5\}$ and $\{m_{\Theta_1}(\theta_1), m_{\Theta_1}(\theta_{123})\} = \{0.8, 0.2\}$, $\{m_{\Theta_2}(\theta_2), m_{\Theta_2}(\theta_{123})\} = \{0.8, 0.2\}$. The results generated by the CFE and other combination rules appear in Table 5.4.

$B \in \mathfrak{F}$	<i>Disj.</i>	<i>Conj.</i>	<i>DCR</i>	<i>YGR</i>	<i>DP1</i>	<i>DP2</i>	<i>PCR</i>	<i>RCR</i>	<i>rCFE</i>	<i>cCFE</i>
\emptyset	-	0.64	-	-	-	-	-	-	-	-
θ_1	-	0.16	0.44	0.16	0.16	0.27	0.48	0.03	0.40	-
θ_2	-	0.16	0.44	0.16	0.16	0.27	0.48	0.03	0.40	-
θ_{12}	0.64	-	-	-	0.64	0.39	-	0.60	-	-
θ_{123}	0.36	0.04	0.12	0.04	0.04	0.07	0.04	0.34	0.20	1.00
Θ	-	-	-	0.64	0.09	-	-	-	-	-
<i>Decision</i>	θ_{12}	\emptyset	θ_{12}	Θ	θ_{12}	θ_{12}	θ_{12}	θ_{12}	θ_{12}	θ_{123}

Table 5.4: Fusion of Contradictory Evidence: Example from [3]

The decision generated by most of the combination rules favor θ_{12} . YGR is different because it basically avoids making a decision. The cCFE narrows down the possibilities to θ_{123} . □

Table 5.5 provides a summary of fusion properties of some of the widely used DST combination rules.

Feature	Dis.	Con.	DCR	YGR	DP	PCR	RCR	CFE
Commutativity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Associativity	Yes	Yes	Yes	No	No	No	Yes	NG
Continuity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Idempotency	No	No	No	No	No	No	No	NG
Non-Exhaustive FoDs	No	No	No	No	No	No	Yes	Yes
Contradictory Evidence	No	Yes	No	No	Yes	Yes	Yes	Yes
Vacuous Fusion	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Evidence Updating	No	No	No	No	No	No	No	Yes
Source/Sensor Weighting	No	No	No	No	No	No	No	Yes
Flexible Parameter Selection	No	No	No	No	No	No	No	Yes

Table 5.5: Summary of Fusion Properties of Commonly Used Combination Rules; Here, Con.=Conjunctive form, Dis.=Disjunctive Form and NG=Not in General.

5.2.2 Evidence Updating

In an updating scenario, even though further refinements are unnecessary, the inertia of available information (i.e., $\mathcal{E}_\Theta[k]$) must be taken into account in the fusion process. CFE can be yield this property by imposing certain restrictions on the parameters.

To show this, consider the CFE-generated $\mathcal{E}_{\Theta_1}[k+1] = \mathcal{E}_{\Theta_1}[k] \bowtie \mathcal{E}_{\Theta_2}[k]$ with $\beta_1(A)[k] = 0, \forall A \neq \Theta_1$, where $\mathcal{E}_{\Theta_i}[k] = \{\Theta_i, \mathfrak{F}_{\Theta_i}[k], m_{\Theta_i}(\cdot)[k]\}, i = 1, 2$. This choice of $\beta_1(\cdot)[k]$ prohibits further refinement of $\mathcal{E}_{\Theta_1}[k]$, because the conditioning operation is performed with only Θ_1 itself.¹² With $\alpha_1[k] = \beta_1(\Theta_1)$, the resulting CFE-generated BoE is exactly what the *Conditional Update Equation* in [52] yields.

¹²Note that $\text{Bl}_{\Theta_1}(\cdot|\Theta_1) \equiv \text{Bl}_{\Theta_1}(\cdot)$; therefore conditioning with respect to $\Theta - 1$ do not change the originally cast beliefs.

Definition 17 (Conditional Update Equation (CUE)). [52] *The belief function associated with the CUE-generated updated BoE $\mathcal{E}_{\Theta_1}[k+1] = \mathcal{E}_{\Theta_1}[k] \triangleleft \mathcal{E}_{\Theta_2}[k]$ is*

$$\text{Bl}_{\Theta_1}(B)[k+1] = \alpha_1[k] \text{Bl}_{\Theta_1}(B)[k] + \sum_{A_2 \in \mathfrak{F}_{\Theta_2}[k]} \beta_2(A_2)[k] \text{Bl}_{\Theta_1}^{(\Theta_2)}(B|A_2)[k], \quad (5.12)$$

whenever, $\text{Pl}_{\Theta_2}(\Theta_1|\Theta_2)[k] > 0$. Here, the non-negative and real parameters satisfy

$$\alpha_1[k] + \sum_{A_2 \in \mathfrak{F}_{\Theta_2}[k]} \beta_2(A_2)[k] = 1, \text{ for all } k. \quad \square$$

5.2.3 Fusion of Multiple BoEs

To fuse multiple BoEs, one can of course repeatedly apply the CFE. But, the fused BoE may depend on the order of fusion, because the CFE does not possess the associativity property, in general. While this may not be a concern in some applications, for instance in evidence updating, certain situations call for aggregation of evidence with the fused BoE being independent of how the evidence sources are ordered for fusion.

5.2.3.1 Order-Independent Fusion for Multiple BoEs

Here we define a multi-BoE extension to the CFE that preserves associativity.

Definition 18. *The CFE-generated fused BoE $\mathcal{E}_{\Theta} = \mathcal{E}_{\Theta_1} \bowtie \mathcal{E}_{\Theta_2} \bowtie \cdots \bowtie \mathcal{E}_{\Theta_n}$ is*

$$\text{Bl}_{\Theta}(B) = \sum_{i=1}^n \sum_{A_i \in \mathfrak{F}_{\Theta_i}} \beta_i(A_i) \text{Bl}_{\Theta}^{(\Theta_i)}(B|A_i), \quad (5.13)$$

whenever $\text{Pl}_{\Theta_i}(\Theta|\Theta_i) > 0$, for $i = 1, \dots, n$. Here, the non-negative real parameters

$$\beta_i(\cdot), i = 1, \dots, n \text{ satisfy } \sum_{i=1}^n \sum_{A_i \in \mathfrak{F}_{\Theta_i}} \beta_i(A_i) = 1. \quad \square$$

Claim 12. *The belief function $\text{Bl}_{\Theta}(\cdot) : 2^{\Theta} \mapsto [0, 1]$ generated by multiple-BoE extension of CFE in Definition 18 is a valid belief function on the FoD Θ . \square*

Proof. Similar to the proof of Claim 9, this follows from the fact that an arbitrary convex combination of belief functions defined on one FoD remains a valid belief function on the same FoD. ■

The plausibility and mass functions associated with the belief function in Definition 18 of multiple-BoE form of CFE can be easily derived as earlier. Hence, we simply state the result without proof.

Claim 13. *The plausibility and mass functions associated with the CFE-generated fused BoE $\mathcal{E}_\Theta = \mathcal{E}_{\Theta_1} \bowtie \mathcal{E}_{\Theta_2} \bowtie \cdots \bowtie \mathcal{E}_{\Theta_n}$ are, respectively,*

$$\text{Pl}_\Theta(B) = \sum_{i=1}^n \sum_{A_i \in \mathfrak{F}_{\Theta_i}} \beta_i(A_i) \text{Pl}_{\Theta}^{(\Theta_i)}(B|A_i); \quad (5.14)$$

$$m_\Theta(B) = \sum_{i=1}^n \sum_{A_i \in \mathfrak{F}_{\Theta_i}} \beta_i(A_i) m_{\Theta}^{(\Theta_i)}(B|A_i); \quad (5.15)$$

whenever $\text{Pl}_{\Theta_i}(\Theta|\Theta_i) > 0$, for $i = 1, 2$. Here, the parameters $\beta_i(\cdot)$, $i = 1, \dots, n$ are non-negative, real and satisfy $\sum_{i=1}^n \sum_{A_i \in \mathfrak{F}_{\Theta_i}} \beta_i(A_i) = 1$. □

5.2.3.2 Order-Independent Iterative-CFE for Multiple BoEs

The strategy in Definition 18 allows for order-independent fusion of multiple BoEs. Here, we provide an iterative strategy, which comes in extremely handy in scenarios that call for order-independent aggregation of multiple BoEs one-at-a-time, e.g., a voting scenario where the evidence has to be aggregated as they “come in.” Let us proceed as follows.

Let $\mathcal{E}_{\{1:i\}} = \{\Theta, \mathfrak{F}_{\{1:i\}}, m_{\{1:i\}}(\cdot)\}$ denote the BoE obtained by iteratively fusing the BoEs \mathcal{E}_{Θ_j} , $j = 1, \dots, i$; hence,

$$\mathcal{E}_{\{1:i\}} = \left(\cdots ((\mathcal{E}_{\Theta_1} \bowtie \mathcal{E}_{\Theta_2}) \bowtie \mathcal{E}_{\Theta_3}) \bowtie \cdots \bowtie \mathcal{E}_{\Theta_i} \right), \quad \text{for } i \geq 2.$$

Now, in order to proceed, note that the task is to iteratively fuse the BoEs as

$$\mathcal{E}_{\{1:i+1\}} = \mathcal{E}_{\{1:i\}} \bowtie \mathcal{E}_{\Theta_{i+1}}, \quad \text{for } i = 1, \dots, n-1,$$

so that $\mathcal{E}_{\{1:n\}} = \mathcal{E}_{\Theta_1} \bowtie \mathcal{E}_{\Theta_2} \bowtie \cdots \bowtie \mathcal{E}_{\Theta_n}$. Here, we take $\mathcal{E}_{\{1:1\}} \equiv \mathcal{E}_{\Theta_1}$.

Step 1—Compute $\mathcal{E}_{\{1:2\}} = \mathcal{E}_{\Theta_1} \bowtie \mathcal{E}_{\Theta_2}$. The corresponding CFE parameters must satisfy

$$\sum_{A_1 \in \mathfrak{F}_{\Theta_1}} \beta_1(A_1) + \sum_{A_2 \in \mathfrak{F}_{\Theta_2}} \beta_2(A_2) = 1. \quad (5.16)$$

Step 2—Compute $\mathcal{E}_{\{1:3\}} = \mathcal{E}_{\{1:2\}} \bowtie \mathcal{E}_{\Theta_3}$. Again, the corresponding CFE parameters must satisfy

$$\sum_{A_{\{1:2\}} \in \mathfrak{F}_{\{1:2\}}} \beta_{\{1:2\}}(A_{\{1:2\}}) + \sum_{A_3 \in \mathfrak{F}_{\Theta_3}} \beta_3(A_3) = 1. \quad (5.17)$$

However, this would entail conditioning in $\mathcal{E}_{\{1:2\}}$, which would not yield the required result $\mathcal{E}_{\{1:n\}}$ at the end. What we need is to select the CFE parameters s.t. the already fused result in $\mathcal{E}_{\{1:2\}}$ is retained. To achieve this objective, invoke the CUE form of the fusion; i.e., perform $\mathcal{E}_{\{1:3\}} = \mathcal{E}_{\{1:2\}} \triangleleft \mathcal{E}_{\Theta_3}$ with CUE parameters satisfying

$$\alpha_{\{1:2\}} + \sum_{A_3 \in \mathfrak{F}_{\Theta_3}} \beta_3(A_3) = 1, \quad \text{with } \alpha_{\{1:2\}} > 0 \quad (5.18)$$

Step i —Compute $\mathcal{E}_{\{1:i+1\}} = \mathcal{E}_{\{1:i\}} \triangleleft \mathcal{E}_{\Theta_{i+1}}$. As in (5.18), we now must choose the CUE parameters to satisfy

$$\alpha_{\{1:i\}} + \sum_{A_{i+1} \in \mathfrak{F}_{\Theta_{i+1}}} \beta_{i+1}(A_{i+1}) = 1, \quad \text{with } \alpha_{\{1:i\}} > 0 \quad (5.19)$$

In this manner, at the conclusion of Step $(n-1)$, we would have computed the required fused result $\mathcal{E}_{\{1:n\}} = \mathcal{E}_{\Theta_1} \bowtie \mathcal{E}_{\Theta_2} \bowtie \cdots \bowtie \mathcal{E}_{\Theta_n}$. Note that, knowledge of the total number of BoEs n that are to be fused is not required. What we have effectively shown is that *a step-wise update using CUE is identical to an associative fusion with CFE*.

Example 8. For a survey that calls for each incoming piece of evidence to be equally ‘weighted’ for fusion, the above order-independent iterative strategy can be used by picking the parameters as

$$\sum_{A_i \in \mathfrak{F}_{\Theta_i}} \beta_i(A_i) = \begin{cases} 1/2, & \text{for } i = 1, 2; \\ 1/i, & \text{for } i > 2; \end{cases}$$

and $\alpha_{\{1:i\}} = i/i+1$, for $i > 2$. □

5.2.3.3 Evidence Updating with Multiple BoEs

To achieve order-independence when a KB is being updated with multiple new pieces of evidence, we can use

Definition 19. The belief function associated with the CUE-generated updated BoE $\mathcal{E}_{\Theta_1}[k+1] = \mathcal{E}_{\Theta_1}[k] \triangleleft (\mathcal{E}_{\Theta_2}[k] \bowtie \cdots \bowtie \mathcal{E}_{\Theta_n}[k])$ is

$$\text{Bl}_{\Theta_1}(B)[k+1] = \alpha_1[k] \text{Bl}_{\Theta_1}(B)[k] + \sum_{i=2}^n \sum_{A_i \in \mathfrak{F}_{\Theta_i}[k]} \beta_i(A_i)[k] \text{Bl}_{\Theta_1}^{(\Theta_i)}(B|A_i)[k], \quad (5.20)$$

whenever $\text{Pl}_{\Theta_i}(\Theta|\Theta_i)[k] > 0$, for $i = 1, \dots, n$. Here, the CUE parameters are non-negative, real and satisfy $\alpha_1[k] + \sum_{i=2}^n \sum_{A_i \in \mathfrak{F}_{\Theta_i}[k]} \beta_i(A_i)[k] = 1$, for all k . □

5.3 Selection of Parameters

With the appropriate parameters being utilized, the CFE provides enormous flexibility to cater to a variety of fusion requirements (evidence fusion and updating, refinement of incoming information to focus on relevant evidence, order-preserving fusion of multiple BoEs, etc.). Any CFE-based fusion of n BoEs \mathcal{E}_{Θ_i} , $i = 1, \dots, n$, to yield the BoE \mathcal{E}_{Θ} requires one to select following parameters:

- (a) FoD Θ for the fused BoE;
- (b) subset of conditioning sets in \mathfrak{F}_{Θ_i} , $i = 1, \dots, n$; and
- (c) combinations weights $\beta_i(A)$, $\forall A \in \mathfrak{F}_{\Theta_i}$, $i = 1, \dots, n$.

How one goes about selecting the appropriate parameter configuration however is highly dependent on the application and domain. Here, we propose a few strategies that are applicable for general fusion and updating scenarios.

5.3.1 Selection of the FoD for fused BoE Θ

The choice of FoD for evidence updating is trivial. However, for an evidence combination, one has the flexibility of picking a relevant FoD with the minimal condition $\Theta \cap \Theta_i \neq \emptyset$, $i = 1, \dots, n$. In addition, CFE does not require the ballooning extensions on the FoD of the BoEs to be fused. For a “meaningful” fusion, one should consider the structure of the evidence and other context/application dependent parameters in the process of selecting an appropriate FoD.

5.3.2 Selection of the Conditioning Sets

Evidence from BoE \mathcal{E}_{Θ_i} is further refined for fusion via conditioning with respect to conditioning events $A \in \mathfrak{F}_{\Theta_i}$, $i = 1, \dots, n$. This refinement can be done w.r.t any conditioning set A as long as $A \in \hat{\mathfrak{F}}_{\Theta_i}$. However, a suitable subset of elements can be chosen by setting $\beta_i(A_i) = 0$ for rest of the conditioning events. We present one such strategy, in the next section, where overlapping focal elements are “merged,” effectively reducing the number of conditioning sets.

5.3.3 Selection of the Linear Combination Weights $\beta_i(\cdot)$

These weights allow one to emphasize/de-emphasize the propositions within each conditioning set A . A relative importance measure of evidence provided by a BoE \mathcal{E}_{Θ_i} (e.g., relative credibility/reliability) can also be incorporated into $\beta_i(\cdot)$ s. For this

purpose, let \mathcal{K}_i denote the normalized relative importance measure associated with BoE \mathcal{E}_{Θ_i} , for $i = 1, \dots, n$.

5.3.3.1 Evidence Combination

In an evidence combination scenario, evidence from the BoE \mathcal{E}_{Θ_i} is “weighted” according to the support \mathcal{E}_{Θ_i} itself has for it. One obvious choice would be:

Definition 20 (Receptive fusion (rCFE)). *The parameters $\beta_i(\cdot)$, $i = 1, \dots, n$ for rCFE fusion is given by*

$$\beta_i(A) = \begin{cases} \mathcal{K}_i m_{\Theta_i}(A), & \text{for } \text{Pl}_{\Theta_i}(\Theta|A) > 0; \\ 0, & \text{otherwise.} \end{cases} \quad (5.21)$$

where, \mathcal{K}_i , $i = 1, \dots, n$ are normalization constants. □

Here, each BoE is conditioned with respect to its focal elements and weighed by the associated mass. An alternative is to condition the evidence based on the groups formed by focal elements. For this, first, construct the set of conditioning sets as

$$\mathcal{D}_i = \left\{ \bigcup_{\substack{j \in J; \\ J \subseteq \mathcal{I}_i}} F_j^{(i)} \mid F_k^{(i)} \cap F_\ell^{(i)} \neq \emptyset, \forall k, \ell \in J \right\}, \quad (5.22)$$

where \mathcal{I}_i is an index set for $\mathfrak{F}_{\Theta_i} \equiv \{F_j^{(i)}\}$, $i = 1, \dots, n$. Then, for each $i = 1, \dots, n$ set the parameters as

$$\beta_i(A) = \begin{cases} \mathcal{K}_i \sum_{B \subseteq A} m_{\Theta_i}(B), & A \in \mathcal{D}_i, \text{Pl}_{\Theta_i}(\Theta|A) > 0; \\ 0, & \text{otherwise.} \end{cases} \quad (5.23)$$

Here, evidence is conditioned on non-overlapping sets formed by taking unions of intersecting focal elements and weighed by their cumulative mass.

Remark:

Often the case is that, in practice, Θ_i is assigned a smaller weight as a way of representing overall uncertainty, not necessarily to support the proposition Θ_i . To account for this, when $\Theta_i \in \mathfrak{F}_{\Theta_i}$, one can obtain \mathcal{D}_i by first forming a \mathcal{D}'_i with $\mathfrak{F}_{\Theta_i} \setminus \Theta_i$ as in (5.22) and then setting $\mathcal{D}_i = \mathcal{D}'_i \cup \Theta_i$, for $i = 1, \dots, n$.

In the case of two BoE fusion ($n = 2$), one can use the following interesting assignment, where evidence from each BoE can be weighted on the support the other BoE has for it.

Definition 21 (Cautious Fusion (cCFE)). *The parameters $\beta_i(\cdot)$, $i = 1, 2$ for cCFE fusion is given by*

$$\beta_i(A) = \begin{cases} \mathcal{K}_i m_{\Theta_j}(A \cap \Theta_i), & \text{for } \text{Pl}_{\Theta_i}(\Theta|A) > 0, i \neq j; \\ 0, & \text{otherwise.} \end{cases} \quad (5.24)$$

where, \mathcal{K}_i , $i = 1, \dots, n$ are normalization constants. □

Note that the conditioning based on grouped focal elements can be easily extended to this case as well.

Example 9. *Consider the fusion scenario in Example 4. Given $\Theta = (bc)$ and the relative importance of the BoEs as $\mathcal{E}_{\Theta_1} : \mathcal{E}_{\Theta_2} = 5 : 3$, let us compute the linear combination weights for rCFE and cCFE cases.*

Case 1. rCFE Parameters: *We have, $\mathcal{K}_1 : \mathcal{K}_2 = 5 : 3$, since $\mathcal{E}_{\Theta_1} : \mathcal{E}_{\Theta_2} = 5 : 3$; now, we get $\beta_1(ab) = 0.4\mathcal{K}_1$, $\beta_1(bc) = 0.2\mathcal{K}_1$, $\beta_2(b) = 0.8\mathcal{K}_2$ and $\beta_2(bc) = 0.2\mathcal{K}_2$. Using $\beta_1(ab) + \beta_1(bc) + \beta_2(b) + \beta_2(bc) = 1$ with $\mathcal{K}_1 : \mathcal{K}_2 = 5 : 3$, we obtain $\mathcal{K}_1 = 5/6$ and $\mathcal{K}_2 = 1/2$. Therefore, $\beta_1(ab) = 1/3$, $\beta_1(bc) = 1/6$, $\beta_2(b) = 2/5$ and $\beta_2(bc) = 1/10$.*

Note that the focal element a in \mathcal{E}_{Θ_1} does not generate any conditioning events, since $\text{Pl}_{\Theta_1}(\Theta|a) = \text{Pl}_{\Theta_2}(\Theta|a) = 0$.

Case 2. cCFE Parameters: Similarly, we get $\beta_1(ab) = 0.8\mathcal{K}_1$, $\beta_1(bc) = 0.2\mathcal{K}_1$, $\beta_1(abc) = 0.2\mathcal{K}_1$ and $\beta_2(bc) = 0.2\mathcal{K}_2$. As before, with $\mathcal{K}_1 : \mathcal{K}_2 = 5 : 3$, we obtain $\mathcal{K}_1 = 25/33$, $\mathcal{K}_2 = 5/11$. Then, $\beta_1(ab) = 20/33$, $\beta_1(bc) = 5/33$, $\beta_1(abc) = 5/33$ and $\beta_2(bc) = 1/11$. \square

5.3.3.2 Evidence Updating

Evidence updating is fundamentally different from a combination operation. In an updating scenario one also has to take into account the inertia of the existing BoE, in addition to weighting the incoming evidence. Consider the updating process $\mathcal{E}_{\Theta_1}[k+1] = \mathcal{E}_{\Theta_1}[k] \triangleleft \mathcal{E}_{\Theta_2}[k] \bowtie \dots \bowtie \mathcal{E}_{\Theta_n}[k]$, $\forall k$ as in Definition 19.

5.3.3.2.1 Selection of $\alpha_1[k]$ The weight $\alpha_1[k]$ in Definition 19 can be interpreted as a measure that indicates the flexibility or inertia of the originally cast evidence. With $\alpha_1[k] = 1$, one can model complete inflexibility of the available evidence towards changes (e.g., when it perceives the incoming evidence to be completely unreliable, when the original BoE is formed from a vast collection of reliable data, etc.) On the other hand, $\alpha_1[k] = 0$ captures the complete flexibility of the available evidence towards changes (e.g., when it perceives the incoming evidence to be completely reliable, when the original BoE has little or no credible knowledge base to begin with, etc.). Non extreme cases can be modeled with an $\alpha_1[k] \in (0, 1)$. For instance, one can set $\alpha_1[k] = M/(M+1)$, where M is the number of “pieces” of evidence on which the available evidence is based upon. This treats each piece of already gathered evidence and the new piece of incoming evidence as having equal inertia.

5.3.3.2.2 Selection of $\beta_i[k]$ We propose the following interesting choices that are inspired by the work in [52, 74, 82].

Receptive Update Strategy (rCUE): This strategy “weights” the incoming evidence from \mathcal{E}_{Θ_i} according to the support \mathcal{E}_{Θ_i} itself has for it. In other words, we are “receptive” to what \mathcal{E}_{Θ_i} thinks.

$$\beta_i(A) = \begin{cases} \mathcal{K}_i m_{\Theta_i}(A), & \text{for } \text{Pl}_{\Theta_i}(\Theta|A) > 0; \\ 0, & \text{otherwise.} \end{cases} \quad (5.25)$$

Cautious Update Strategy (cCUE): This strategy “weights” the incoming evidence from \mathcal{E}_{Θ_i} according to the support \mathcal{E}_{Θ_1} has for it. In other words, being “cautious,” we are checking if incoming sources are agreeable with currently available evidence.

$$\beta_i(A) = \begin{cases} \mathcal{K}_i m_{\Theta_1}(A \cap \Theta_1), & \text{for } \text{Pl}_{\Theta_i}(\Theta|A) > 0; \\ 0, & \text{otherwise.} \end{cases} \quad (5.26)$$

Let us illustrate these with an example.

Example 10. Given a BoE $\mathcal{E}_{\Theta_0}[k] \equiv \{\Theta_0, \mathfrak{F}_{\Theta_0}[k], m_{\Theta_0}[k]\}$, where $\Theta_0 = \{b, c\}$ and mass assignment: $m_{\Theta_0}(c)[k] = 0.2$ and $m_{\Theta_0}(bc)[k] = 0.8$. We would like to update the BoE $\mathcal{E}_{\Theta_0}[k]$ with the evidence given by \mathcal{E}_{Θ_1} and \mathcal{E}_{Θ_2} at time index k as given in Example 3.

BoE $\mathcal{E}_{\Theta_0}[k]$ is fairly new. Hence, we would like to give more “weight” to incoming evidence. So, set $\alpha_0[k]$ to a lower value, say $\alpha_0[k] = 1/5$. Since $\mathcal{E}_{\Theta_0}[k]$ is receptive to incoming evidence, rCUE is more appropriate. First, compute the ported conditional belief values (see Table 5.2). With $\alpha_0[k] = 1/5$, we get $\beta_1(ab)[k] = 0.2$, $\beta_1(bc)[k] = 0.1$, $\beta_2(b)[k] = 0.4$ and $\beta_2(bc)[k] = 0.1$. Then, use Equation (5.20) to compute the updated belief as, $\text{Bl}_{\Theta_0}(b) = 0.68$, $\text{Bl}_{\Theta_0}(c) = 0.1$ and $\text{Bl}_{\Theta_0}(bc) = 1.0$. \square

5.4 An Illustrative Example

In this section, we utilize an extended version of the example scenario in [52] to highlight the various features of the CFE and also to provide insight into parameter selection strategies.

5.4.1 Scenario

A remote military facility uses a suite of hard sensors, viz., \mathcal{S}_H and soft sensors \mathcal{S}_{S_1} , \mathcal{S}_{S_2} and \mathcal{S}_{S_3} , for the identification of objects crossing its perimeter. The objects crossing the perimeter are classified into one of four classes:

$$S \equiv \text{Soldier}, F \equiv \text{Fighter-Jet}, T \equiv \text{Tank}, O \equiv \text{Other};$$

each class, except O which accounts for an object that cannot be sub-classified (e.g., an animal), may further be sub-classified as $f = \text{friendly}$ or $e = \text{enemy}$; so, the exhaustive set of objects of interest is

$$\Theta_{Obj} = \{\underbrace{S_f, S_e}_{\equiv S}, \underbrace{F_f, F_e}_{\equiv F}, \underbrace{T_f, T_e}_{\equiv T}, O\}.$$

5.4.2 Evidence

The BoEs associated with each evidence source is as follows:

Hard Sensor Suite: Sensors belonging to the suite \mathcal{S}_H can identify ground objects, but cannot differentiate between $f = \text{friendly}$ and $e = \text{enemy}$ forces. The corresponding BoE is $\mathcal{E}_{\Theta_H} = \{\Theta_H, \mathfrak{F}_{\Theta_H}, m_{\Theta_H}(\cdot)\}$, where $\Theta_H = \{S, T, O\}$.

Soft Sensors: The soft sensor \mathcal{S}_{S_1} provides the prevailing threat level (TL) in the proximity of the security zone. We assume that this assessment is provided by the base commander, hence a highly reliable source. The soft sensors \mathcal{S}_{S_2} and \mathcal{S}_{S_3} correspond to two cooperating human witnesses. We use \mathcal{S}_{S_1} for refining the

evidence provided by these human witnesses. The BoE corresponding to source \mathcal{S}_{S_i} is taken as $\mathcal{E}_{\Theta_{S_i}} = \{\Theta_{S_i}, \mathfrak{F}_{\Theta_{S_i}}, m_{\Theta_{S_i}}(\cdot)\}$ for $i = 2, 3$, where $\Theta_S = \{S, T, F\}$.

We partition the time axis into intervals, so that the evidence being received within each interval is unchanged. For our simulations, we use five such intervals, (a), (b), (c), (d) and (e). We will use an index k , which may take the values $\{a, b, c, d, e\}$, to identify the interval. For convenience, we will then use the indices $k - 1$ and $k + 1$ to identify the previous and next intervals of interval k . For example, $\mathcal{E}_{\Theta_H}[b]$ is the hard evidence generated within interval (b); $\mathcal{E}_{\Theta_H}[k - 1]$ then identifies the hard evidence generated in interval (a). Table 5.6 shows the evidence being generated by the sources \mathcal{S}_H and \mathcal{S}_{S_i} , $i = 1, 2, 3$, within each interval and their corresponding DS theoretic mass assignments.

Evidence Source	Interval (a)	Interval (b)	Interval (c)	Interval (d)	Interval (e)
\mathcal{S}_H : Hard Sensor \mathcal{E}_{Θ_H}	No data $\{\Theta_H\}=\{1.0\}$	$\{T, \Theta_H\}=\{.70, .30\}$	$\{T, \Theta_H\}=\{.70, .30\}$	No data $\{\Theta_H\}=\{1.0\}$	$\{S, O\}=\{.80, .20\}$
\mathcal{S}_{S_1} : Threat Level $\mathcal{A}_i, i=1, 2$	<i>Normal</i> \emptyset	<i>Ground Attack</i> (S, T)	<i>Ground Attack</i> (S, T)	Unknown Θ_*	<i>Ground/Air Attack</i> Θ_S
\mathcal{S}_{S_2} : Wit. Report 1 $\mathcal{E}_{\Theta_{S_2}}$	$S, 90\%$ $\{S, \Theta_S\}=\{.90, .10\}$	$T, 80\%$ $\{T, \Theta_S\}=\{.80, .20\}$	$T, 75\%$ $\{T, \Theta_S\}=\{.75, .25\}$	T or $F, 90\%$ $\{(T, F), \Theta_S\}=\{.90, .10\}$	$S, 90\%$ $\{S, \Theta_S\}=\{.90, .10\}$
\mathcal{S}_{S_3} : Wit. Report 2 $\mathcal{E}_{\Theta_{S_3}}$	No data $\{\Theta_S\}=\{1.0\}$	$F, 70\%$ $\{F, \Theta_S\}=\{.70, .30\}$	$F, 75\%$ $\{F, \Theta_S\}=\{.75, .25\}$	$T, 90\%$ $\{T, \Theta_S\}=\{.90, .10\}$	$F, 80\%$ $\{F, \Theta_S\}=\{.80, .20\}$

Table 5.6: Evidence Gathered from Soft and Hard Sensor Suites

5.4.3 Evidence Updating/Fusion

In our simulations, we look at three schemes, FS1, FS2 and FS3, that a fusion center may employ to maintain a KB. Let us denote the BoE generated at this fusion center as $\mathcal{E}_{\Theta_*}[k] = \{\Theta_*, \mathfrak{F}_{\Theta_*}[k], m_{\Theta_*}(\cdot)[k]\}$, where Θ_* is the FoD being retained at the

KB. Corresponding to each fusion scheme, the fused BoEs obtained in the intervals (a)-(e) are tabulated in Table 5.7.

Fusion Scheme	$B \in \mathfrak{F}_{\Theta_*}[k] \downarrow$	$m_{\Theta_*}(\bullet)[k]$				
		$k = a$	$k = b$	$k = c$	$k = d$	$k = e$
FS1	Θ_H	1.0000	0.3000	0.7000	0.7000	0.1205
	T	-	0.7000	0.3000	0.3000	0.5915
	S	-	-	-	-	0.2880
FS2	Θ_{Obj}	1.0000	-	-	-	-
	Θ_S	-	-	-	0.1000	0.0750
	Θ_H	-	0.1500	0.1500	-	-
	S, T	-	0.1000	0.1000	-	-
	F, T	-	-	-	0.4500	-
	T	-	0.7500	0.7500	0.4500	-
	S	-	-	-	-	0.6250
	F	-	-	-	-	0.2000
	O	-	-	-	-	0.1000
FS3	Θ_{Obj}	1.0000	-	-	-	-
	Θ_S	-	0.1250	0.1250	0.1225	0.0872
	Θ_H	-	0.1500	0.1500	0.1350	0.0135
	F, T	-	-	-	0.0450	0.0045
	T	-	0.5500	0.5500	0.5400	0.0540
	S	-	-	-	-	0.5450
	F	-	0.1750	0.1750	0.1575	0.2158
	O	-	-	-	-	0.0800

Table 5.7: Fused Results Corresponding to Example 5.4

5.4.3.1 FS1: Updating Hard Evidence

Here, the fused BoE is given by $\mathcal{E}_{\Theta_*}[k+1] = \mathcal{E}_{\Theta_*}[k] \triangleleft \mathcal{E}_{\Theta_H}[k+1]$, where $\Theta_* = \Theta_H$ with $\beta_H(\Theta_H)[k] = 1 - \alpha_H[k]$. This simulates a scenario where soft information is unavailable and only the hard evidence is used without any refinement.

Note that, $\mathcal{E}_{\Theta_*}[a]$ being vacuous, we set $\alpha_H[b] = 0$ for the update in interval (b).

For $k = c$, we set $\mathcal{E}_{\Theta_*}[k] = \mathcal{E}_{\Theta_*}[k - 1]$ because there is no change in the evidence provided by \mathcal{S}_H . In interval (d), with no incoming evidence, we set $\alpha_H[d] = 1$ to preserve the existing KB. In interval (e), the KB is updated with $\alpha_H[e] = 0.7$. The fused results in each interval appear in Fig. 5.1.

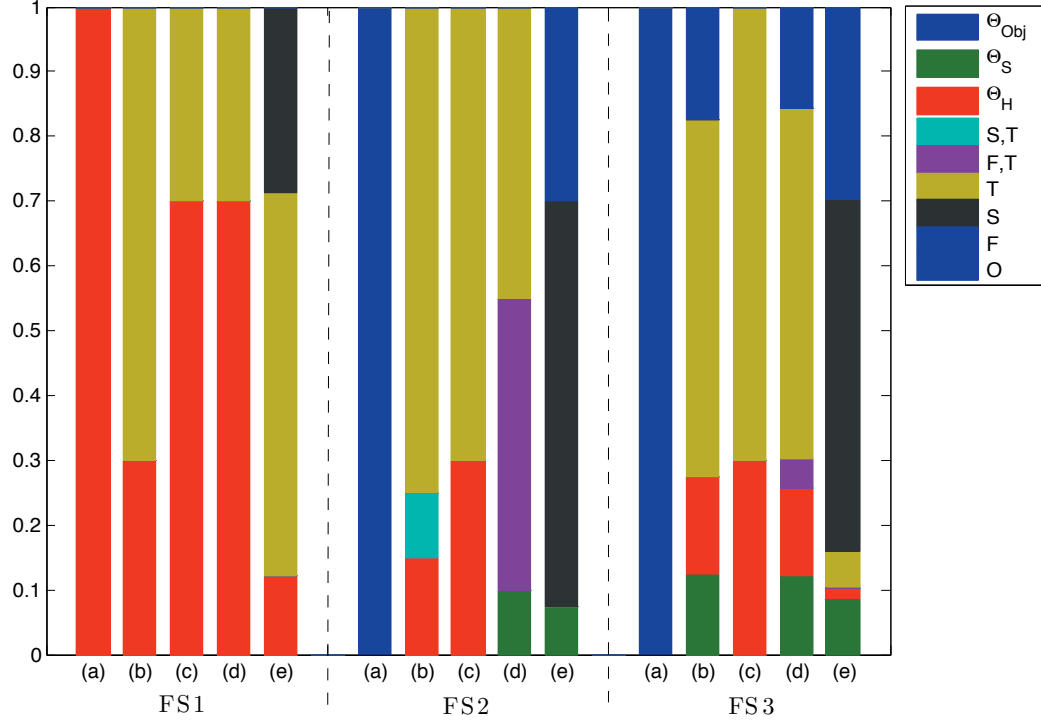


Figure 5.1: Fusion results of Example 5.4. Fused mass assignments in each interval is plotted in its own section corresponding to the fusion scheme. For instance, $m_{\Theta_*}(\cdot)[b]$ for FS1 is plotted in Section FS1 under (b).

5.4.3.2 FS2: Fusion of Soft/Hard Evidence

Here, hard evidence in each interval is fused with soft information from the same interval. We refine the incoming soft information from the human witnesses \mathcal{S}_{S_2} and \mathcal{S}_{S_3} using the threat level information (TL) from \mathcal{S}_{S_1} . The hard evidence from \mathcal{S}_H is used ‘as is’ (i.e., conditioning on Θ_H). Thus, the fused BoE is given by $\mathcal{E}_{\Theta_*}[k] =$

$\mathcal{E}_{\Theta_H}[k] \triangleleft (\mathcal{E}_{\Theta_{S_2}}[k] \bowtie \mathcal{E}_{\Theta_{S_3}}[k])$. We use $\Theta_* = \Theta_{Obj}$.

In interval (a), soft information is “filtered out,” because $TL = Normal$. In interval (b), evidence from \mathcal{S}_3 is filtered out because the threat level indicates a “Ground Attack.” Fusion is performed with $\alpha_H[b] = \beta_2(S, T)[b] = 0.5$, which puts equal weight to \mathcal{E}_{Θ_H} and $\mathcal{E}_{\Theta_{S_2}}$. Interval (c) can be handled similarly. But we can exploit the continuity property of the CFE and not update the KB because the evidence change in $\mathcal{E}_{\Theta_{S_2}}$ is minimal. In real-life applications, such observations can be used to cut down on computational and energy costs (e.g., in distributed sensor networks). So, we refrain from updating the KB in this interval. Interval (d) illustrates a scenario, where hard sensors fail to capture evidence and domain experts (in this example, the base commander) have not yet been able to assess the situation. In this case, soft information is fused with no further refinement. We use $\beta_i(\Theta_*)[d] = 0.5, i = 2, 3$, to equally weigh the evidence from \mathcal{S}_{S_2} and \mathcal{S}_{S_3} . Fusion in interval (e) is similar to interval (b) and we pick the parameters $\alpha_H[e] = 0.5, \beta_2(\Theta_S)[e] = \beta_3(\Theta_S)[e] = 0.25$. The fused results in each interval appear in Fig. 5.1.

5.4.3.3 FS3: Updating Hard Evidence with Soft Evidence

Here, we consider the ‘complete’ fusion scenario where the KB is updated by evidence from both soft and hard sensors. The fused BoE is given by $\mathcal{E}_{\Theta_*}[k+1] = \mathcal{E}_{\Theta_*}[k] \triangleleft (\mathcal{E}_{\Theta_H}[k] \bowtie \mathcal{E}_{\Theta_{S_2}}[k] \bowtie \mathcal{E}_{\Theta_{S_3}}[k])$, where $\Theta_* = \Theta_{Obj}$. As above, we refine the soft information using TL and use hard evidence as is.

Interval (a) remains vacuous as before because $TL = Normal$. In interval (b), we pick $\alpha_*[b] = 0, \beta_H(\Theta_H)[b] = 0.5$ and $\beta_i(S, T)[b] = 0.25, i = 2, 3$, thus giving more weight to the hard data. In interval (c), since the change in evidence provided by the sensors is marginal, as implied by the continuity property of the CFE, $\mathcal{E}_{\Theta_H}[c] \bowtie \mathcal{E}_{\Theta_{S_2}}[c] \bowtie \mathcal{E}_{\Theta_{S_3}}[c] \approx \mathcal{E}_{\Theta_*}[b]$ (with the same parameters). Thus, we refrain from updating the KB. However, in interval (d), lack of new evidence forces the KB

to preserve its existing evidence. Thus, we set $\alpha_*[d] = 0.9$ and update from the soft data with $\beta_i(\Theta_*)[d] = 0.05$, $i = 2, 3$. In interval (e), both TL and hard sensor information are available. Thus, in order to take more recent and reliable information into account, evidence update is carried out with the KB allowed to be more flexible to change. As usual, we also refine soft evidence using TL . Accordingly, we set the parameters as $\alpha_*[e] = 0.1$, $\beta_H(\Theta_H)[e] = 0.4$ and $\beta_i(\Theta_S)[e] = 0.25$, $i = 2, 3$. The fused results in each interval appears in Fig. 5.1.

5.4.4 Analysis

As this example illustrates, the CFE parameters can be used to easily account for relative reliability of evidence sources. The conditioning operation allows one to refine the incoming evidence. For instance, in interval (b) of the FS2 scheme, when $TL = \textit{Ground Attack}$, the evidence provided by \mathcal{S}_{S_3} was ignored by the CFE and only the relevant evidence from \mathcal{S}_{S_2} was taken. This is a very useful feature in soft/hard data fusion, where sources often provide irrelevant information.

5.5 Application Examples

In this section, we illustrate the application of CFE to real life evidence combination and updating scenarios. Each application also shows basic steps of DST evidence modeling.

5.5.1 NLP Application

Here, we illustrate how the CFE can be used for the task of incorporating soft evidence in the form of text statements, into the fusion task. The issues of Natural Language Processing (NLP) parsing of text, logical form extraction, and conversion to DS theoretic forms are not addressed here. We use this application example to illustrate how the CFE can be used for the pertinent fusion tasks.

5.5.1.1 Scenario

A suspicious activity in the proximity of a military base was reported. Reports from the various hard sensors (e.g., metal, magnetic, IR) deployed around the perimeter of the base confirms vehicle activity; and analysis of night vision cameras confines the set of possible vehicles to one in $\Theta_V = \{Jeep, Truck, Car\} \equiv \{Jp, Tk, Cr\}$. The task of the base commander is to determine the most probable suspect and the vehicle driven by the suspect.

5.5.1.2 Setup

The base commander maintains a “blacklisted” group of personnel in $\Theta_\pi = \{Andy, Bob, Ken, Larry\} \equiv \{A, B, K, L\}$, from which he usually picks the initial suspects. The commander gets soft evidence from two human witnesses WS_1 and WS_2 and also from a public database DB_3 (see Table 5.8) containing demographic information on $\Theta_S \times \Theta_\Pi$, where $\Theta_S = \{Tall, Med, Short\} \equiv \{Tl, Md, St\}$ and $\Theta_\Pi = \{Andy, Bob, Chuck, \dots, Jude\} \equiv \{A, B, C, \dots, J\}$.

Source	Evidence	Confidence
WS_1	“A tall man was driving a truck or jeep”	[0.7]
WS_2	“Andy drives a truck”	[0.9]
DB_3	$\langle \text{Height} = \text{Tall} \rangle \implies \langle \text{Person} = \text{Bob} \rangle$	[0.5, 0.8]

Table 5.8: Evidence gathered from WS_1 , WS_2 and DB_3 .

The values in square brackets indicate the confidence each source places on its own evidence. Table 5.9 shows the reliability and DS theoretic evidence model corresponding to each evidence source. Note that $|\Theta_\Pi| = 10$, $\Theta_\pi \not\subset \Theta_\Pi$ and $\Theta_\Pi \cap \Theta_\pi = \{A, B\}$. The CFE’s ability to handle non-exhaustive frames without having to expand the FoDs using computationally expensive ballooning extensions becomes very handy in this situation.

Source	Source Reliability	DS Model	FoD
WS_1	r_1	\mathcal{E}_{Θ_1}	$\Theta_1 = \Theta_S \times \Theta_V \times \Theta_\pi$
WS_2	r_2	\mathcal{E}_{Θ_2}	$\Theta_2 = \Theta_S \times \Theta_V \times \Theta_\pi$
DB_3	r_3	\mathcal{E}_{Θ_3}	$\Theta_3 = \Theta_S \times \Theta_V \times \Theta_\Pi$

Table 5.9: Evidence models of WS_1 , WS_2 and DB_3 .

5.5.1.3 Modeling

Let us use the following models for the soft evidence provided by witnesses and the database:

$$\begin{aligned}
\mathcal{E}_{\Theta_1} : m_{\Theta_1}(Tl \times (Jp, Tk) \times \Theta_\pi) &= 0.7r_1 \\
m_{\Theta_1}(\Theta_S \times \Theta_V \times \Theta_\pi) &= 1 - 0.7r_1 \\
\mathcal{E}_{\Theta_2} : m_{\Theta_2}(\Theta_S \times Tk \times A) &= 0.9r_2 \\
m_{\Theta_2}(\Theta_S \times \Theta_V \times \Theta_\pi) &= 1 - 0.9r_2 \\
\mathcal{E}_{\Theta_3} : m_{\Theta_3}(\Theta_3(Tl \times \Theta_V \times B) &= r_3c_1 \\
m_{\Theta_3}((Md, St) \times \Theta_V \times \neg B) &= r_3(1 - c_2) \\
m_{\Theta_3}(\Theta_S \times \Theta_V \times \Theta_\Pi) &= 1 - r_3(1 + c_1 - c_2)
\end{aligned}$$

Remarks:

1. We use simple, intuitive models to capture the soft evidence from witnesses WS_1 and WS_2 . The reliability associated with an evidence source is incorporated by simply discounting the initial mass assignments. For instance, if the reliability r_1 of WS_1 is very low, we may want to give a lower weight to the proposition $Tl \times (Jp, Tk) \times \Theta_\pi$.
2. Textual information or expert opinions are often modeled as logical implication rules. The ability to combine such evidence into hard evidence is of significant importance, especially in military, medical, and other sensitive domains.

5.5.1.4 Evidence Updating with CFE

We use the following evidence updating strategy (see Fig. 5.2) to fuse soft/hard evidence. Let the BoE, $BoE^{(k)} \equiv \mathcal{E}_\Theta[k] = \{\Theta_S \times \Theta_V \times \Theta_\pi, \mathfrak{F}_\Theta[k], m_\Theta(\cdot)[k]\}$ be the evidence BoE at the k -th update. We initialize BoE $\mathcal{E}_\Theta[0]$ with a vacuous BoE representing the hard evidence, “The observed vehicle is in Θ_V ”. At the k -th update cycle, we compute the update $\mathcal{E}_\Theta[k] = \mathcal{E}_\Theta[k-1] \bowtie \mathcal{E}_k$, $k = 1, \dots, 3$. The idea here is to refine (i.e., update) the evidence obtained from hard sensors with the soft evidence from the witnesses and the public database in order to narrow down the possible suspects.

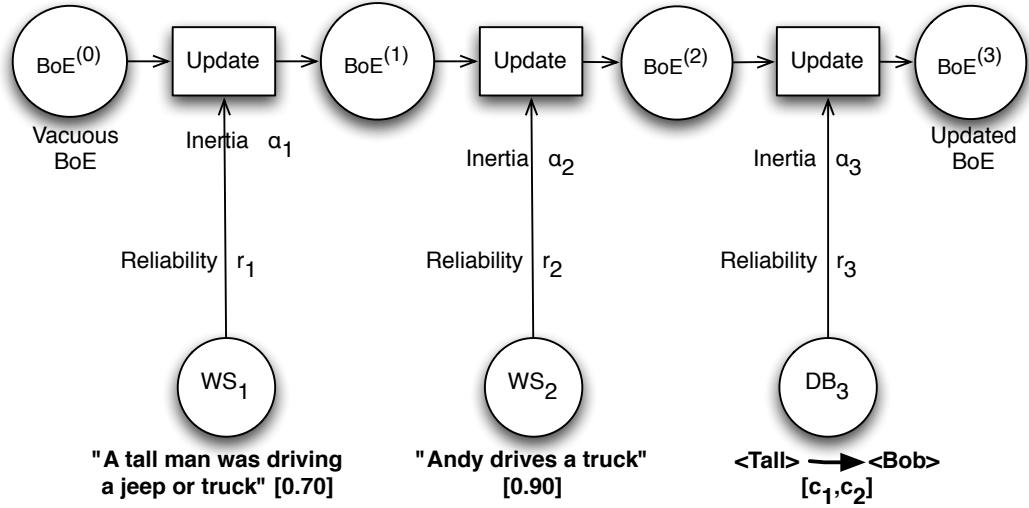


Figure 5.2: Updating scheme used in the example

5.5.1.5 Fusion Results

We obtain the following BoE after updating the initial BoE with all three pieces of evidence:

$$m_\Theta(Tl \times (Jp, Tk) \times \Theta_\pi) = \alpha_3 \alpha_2 (1 - \alpha_1) (0.7 r_1) (2 - 0.7 r_1)$$

$$m_\Theta(\Theta_S \times Tk \times A) = \alpha_3 (1 - \alpha_2) (0.9 r_2) (2 - 0.9 r_2)$$

$$\begin{aligned}
m_{\Theta}(\Theta_S \times \Theta_V \times \Theta_{\pi}) &= \alpha_3 \alpha_2 [\alpha_1 + (1 - \alpha_1)(1 - 0.7r_1)^2] + \alpha_3(1 - \alpha_2)(1 - 0.9r_2)^2 \\
m_{\Theta}(Tl \times \Theta_V \times B) &= r_3 c_1(1 - \alpha_3)[2 - r_3(1 + c_1 - c_2)] \\
m_{\Theta}((Md, St) \times \Theta_V \times A) &= r_3^2(1 - \alpha_3)(c_2 - 1)(1 + c_1 - c_2) + 2r_3(1 - \alpha_3)(1 - c_2) \\
m_{\Theta}(\Theta_S \times \Theta_V \times (A, B)) &= (1 - \alpha_3)[1 - r_3(1 + c_1 - c_2)]^2.
\end{aligned}$$

Note the dependency of these final masses on the source reliability and “inertia,” viz., α_i . Table 5.10 contains the final fusion results for several sets of parameters.

Rule confidence low: $[c_1, c_2] = [0.1, 0.4]$						
$WS_1 \rightarrow$	$r_1 = 0.10$		$r_1 = 0.90$		$r_1 = 0.90$	
$WS_2 \rightarrow$	$r_2 = 0.90$		$r_2 = 0.10$		$r_2 = 0.90$	
Proposition ↓	$BetP_{\Theta}(\cdot)$	$[Bl_{\Theta}(\cdot), Pl_{\Theta}(\cdot)]$	$BetP_{\Theta}(\cdot)$	$[Bl_{\Theta}(\cdot), Pl_{\Theta}(\cdot)]$	$BetP_{\Theta}(\cdot)$	$[Bl_{\Theta}(\cdot), Pl_{\Theta}(\cdot)]$
Tk, A	0.42	[0.30, 0.70]	0.17	[0.01, 0.99]	0.49	[0.42, 0.58]
Jp, A	0.11	[0.00, 0.66]	0.16	[0.00, 0.96]	0.07	[0.00, 0.58]
Cr, A	0.11	[0.00, 0.65]	0.10	[0.00, 0.44]	0.02	[0.00, 0.17]
Tk, B	0.05	[0.00, 0.47]	0.10	[0.00, 0.77]	0.07	[0.00, 0.58]
Jp, B	0.05	[0.00, 0.47]	0.10	[0.00, 0.77]	0.07	[0.00, 0.58]
Cr, B	0.05	[0.00, 0.47]	0.03	[0.00, 0.26]	0.02	[0.00, 0.16]
$(Jp, Tk), A$	0.53	[0.30, 0.96]	0.33	[0.01, 0.99]	0.56	[0.42, 1.00]
$(Jp, Tk), B$	0.10	[0.00, 0.47]	0.20	[0.00, 0.77]	0.14	[0.00, 0.58]
$(Jp, Tk), K$	0.07	[0.00, 0.40]	0.16	[0.00, 0.69]	0.13	[0.00, 0.55]

Rule confidence high: $[c_1, c_2] = [0.6, 0.9]$						
$WS_1 \rightarrow$	$r_1 = 0.1$		$r_1 = 0.9$		$r_1 = 0.9$	
$WS_2 \rightarrow$	$r_2 = 0.1$		$r_2 = 0.1$		$r_2 = 0.9$	
Proposition ↓	$BetP_{\Theta}(\cdot)$	$[Bl_{\Theta}(\cdot), Pl_{\Theta}(\cdot)]$	$BetP_{\Theta}(\cdot)$	$[Bl_{\Theta}(\cdot), Pl_{\Theta}(\cdot)]$	$BetP_{\Theta}(\cdot)$	$[Bl_{\Theta}(\cdot), Pl_{\Theta}(\cdot)]$
Tk, A	0.08	[0.01, 0.99]	0.10	[0.01, 0.99]	0.37	[0.30, 0.70]
Jp, A	0.08	[0.00, 0.77]	0.10	[0.00, 0.77]	0.06	[0.00, 0.47]
Cr, A	0.08	[0.00, 0.76]	0.03	[0.00, 0.26]	0.03	[0.00, 0.18]
Tk, B	0.14	[0.00, 0.96]	0.16	[0.00, 0.96]	0.13	[0.00, 0.66]
Jp, B	0.14	[0.00, 0.96]	0.16	[0.00, 0.96]	0.13	[0.00, 0.66]
Cr, B	0.14	[0.00, 0.95]	0.10	[0.00, 0.44]	0.09	[0.00, 0.36]
$(Jp, Tk), A$	0.16	[0.01, 0.99]	0.20	[0.01, 0.99]	0.43	[0.30, 0.78]
$(Jp, Tk), B$	0.28	[0.00, 0.96]	0.32	[0.00, 0.96]	0.25	[0.00, 0.66]
$(Jp, Tk), K$	0.12	[0.00, 0.69]	0.16	[0.00, 0.69]	0.09	[0.00, 0.40]

Table 5.10: Final fusion results.

5.5.1.6 Analysis of Fusion Results

We have taken DB_3 to be very reliable with $r_3 = 0.9$; r_1 and r_2 denote the reliabilities of WS_1 and WS_2 , respectively. $BetP_\Theta(\cdot)$ column depicts the pignistic probability [67]; the $[Bl_\Theta(\cdot), Pl_\Theta(\cdot)]$ values depict the corresponding belief and plausibility values which can be interpreted as indicating the uncertainty associated with the underlying probability.

The main observation one should make here is that, while one may reach the same conclusion under different circumstances, the uncertainty associated with the decision may vary significantly. For instance, when rule confidence is low, while both scenarios $\{r_1 = 0.1, r_2 = 0.9\}$ and $\{r_1 = 0.9, r_2 = 0.9\}$ favor Andy driving the truck, the uncertainty associated with the latter is much smaller because of the higher reliability of WS_1 . When rule confidence is high, a decision favoring Bob driving has to be made with care because the associated uncertainty is very high. This is one main advantage of DS theory: one can make a decision with a better awareness of the associated uncertainties.

5.5.2 Underwater Objects Characterization

Characterization, management and remediation of military munitions, especially in underwater environments, is a challenging task given all the technical and physical barriers. Optical cameras are better suited for identifying the physical shape of objects. But in underwater, low visibility almost prohibits the use of these cameras. Acoustic imaging is a good alternative to this, but the characteristics of imaging along with numerous artifacts of physical systems which are not easy to model, makes the object recognition task non-trivial. Here, we illustrate the use of CFE for the task at hand by counteracting data imperfections via fusing evidence from multiple perspectives (hence, sources). The data used in the experiments were obtained at a test site located in the Florida Atlantic University premises.

5.5.2.1 Problem Formation

Let $\Theta = \{\theta_1, \dots, \theta_n\}$ be the set of targets of interest. Once an object is observed, optical sensor provides the BoE $\mathcal{E}_{\Theta_O} = \{\Theta, \mathfrak{F}_{\Theta_O}, m_{\Theta_O}(\cdot)\}$ which models the optical evidence. Similarly, sonar sensor provides the BoE $\mathcal{E}_{\Theta_S} = \{\Theta, \mathfrak{F}_{\Theta_S}, m_{\Theta_S}(\cdot)\}$. The objective is to obtain the fused BoE $\mathcal{E}_{\Theta} = \{\Theta, \mathfrak{F}_{\Theta}, m_{\Theta}(\cdot)\}$ to characterize the object observed. We seek for a method to obtain an ‘optimal’ \mathcal{E}_{Θ} for the task at hand by exploring different possible ways of fusing available evidence.

5.5.2.2 Experimental Setup

An object O_{Θ} is placed in the middle of a pool with a known depth and dimensions. An equipment setup consisting of (i) a 1M-pixel digital optical stereo camera, and (ii) a DIDSON sonar video camera operating at 1.1/1.8MHz, is rotated around the target. The height of the equipment setup from the ground plane and the distance to O_{Θ} are recorded. There are 5 possible target types, identified as $\theta_i \neq \theta_6, i = 1, \dots, 5$ where θ_6 denotes any other object.

Target	Description
θ_1	Cylindrical target with length l and radius $h/2$
θ_2	Cylindrical target with length l and radius h
θ_3	Cylindrical target with length $l/2$ and radius $h/2$
θ_4	Spherical target with radius $h/2$
θ_5	Spherical target with radius h

Table 5.11: Target Characteristics

Let α denote the incident angle of sensors to the target. Given α , target dimensions and distance to target from cameras and ground level, the acoustic shadow image $\mathcal{S}h_{\theta_i}(\alpha_{in})$ corresponding to each target type θ_i is computed. To characterize the object O_{Θ} , both optical and acoustic images of the object were taken at incident angles of 0, 30, 60, 90, 120, 150 degrees.

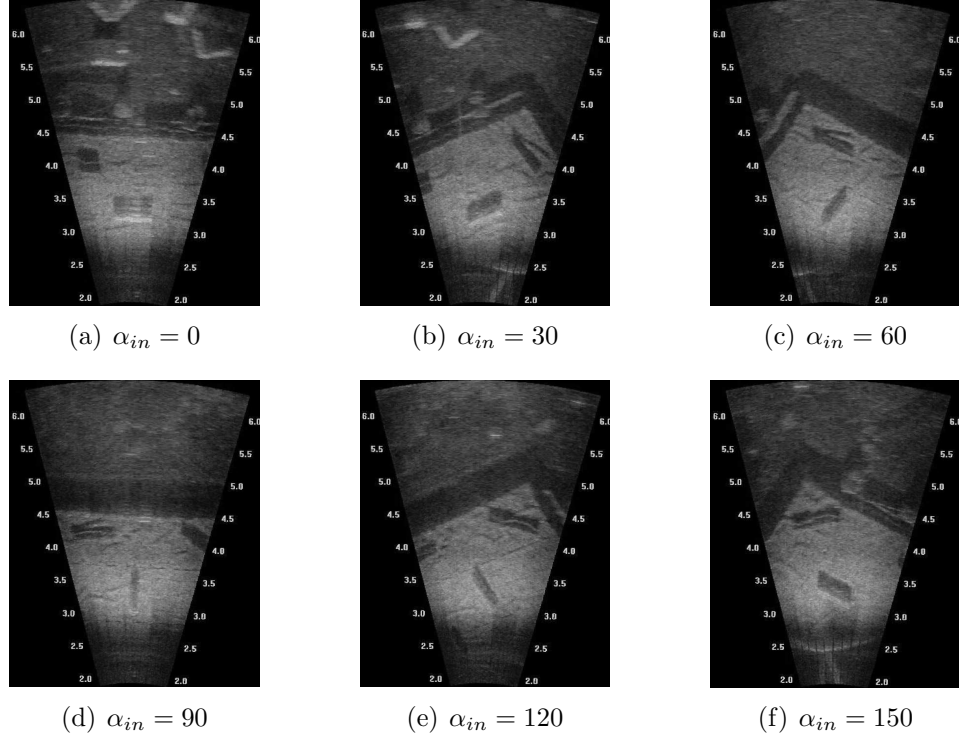


Figure 5.3: Acoustic images of the object O_Θ at angles $0 - 150$ at steps of 30 degrees. Observe the variation of the shadow geometry with α_{in} .

5.5.2.3 Evidence Modeling

We generate \mathcal{E}_O for each α_{in} by manually assigning the weights depending on the ability to visually identify targets. Note that, for simulating a human observer, only the object shapes can be identified. Inability to identify objects precisely, i.e., the uncertainty in classification, in a given view is modeled by assigning a mass to the complete ambiguity Θ .

Shadow analysis on acoustic images allows one to quantify evidence and generate confidence values to represent whether an observed object is of type $\theta_i \in \Theta$. This is done via a measure $\mu_{\theta_i}(\cdot) : \mathcal{Sh}_{\theta_i}(\cdot) \times \mathcal{Sh}_{O_\Theta}(\cdot) \mapsto [0, 1]$, which captures the ‘similarity’ of $\mathcal{Sh}_{\theta_i}(\cdot)$ to $\mathcal{Sh}_{O_\Theta}(\cdot)$, i.e., the shadow image of the object under investigation. We assign

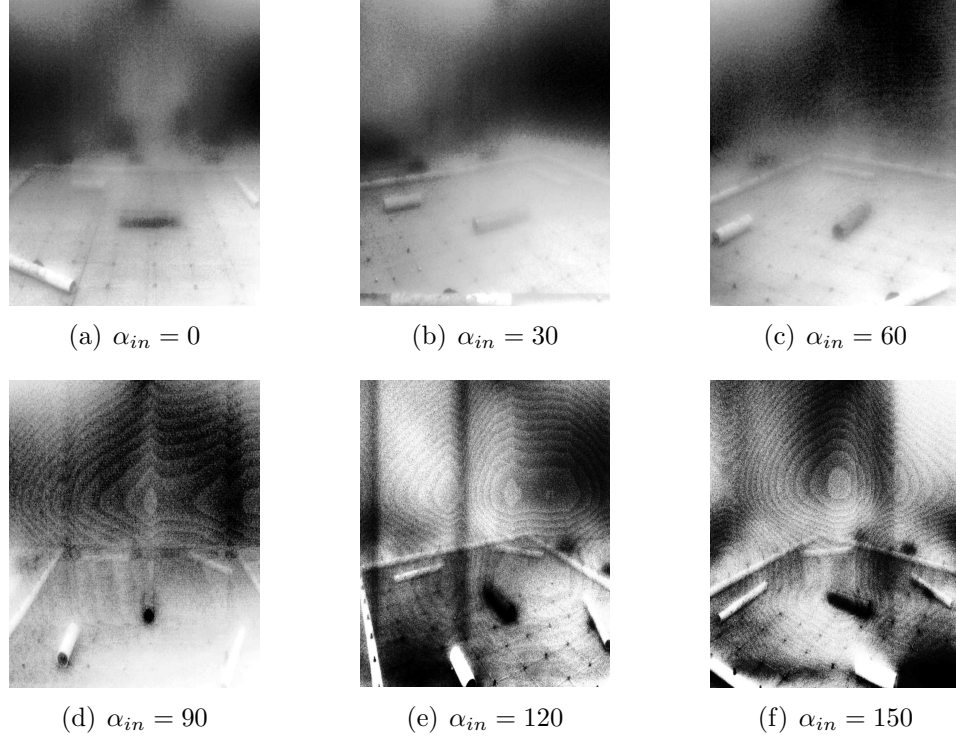


Figure 5.4: Preprocessed optical images of the object O_Θ at angles $0 - 150$ at steps of 30 degrees. In views 5.4(b)-5.4(c), it is clear that object is cylindrical. However, view 5.4(a) could well be some rectangular object. This uncertainty is even more pronounced in view 5.4(d).

the ‘singleton’ mass $m_{\Theta_S}(\theta_i) = 0.2 \times \mu_{\theta_i}(\alpha_{in})$, for each $\theta_i \in \Theta$, and the remaining total mass to $m_{\Theta_S}(\Theta)$, for each α_{in} . See Table 5.12 for an example mass assignment for object O_Θ .

5.5.2.4 Combining Evidence Sources

In the current setup, evidence fusion is not trivial. One can choose to combine/update evidence from individual sources over different α_{in} , or one can combine two different sources for a fixed α_{in} . We take a more intuitive approach to this with the help of CFE.

α_{in}	$\mathcal{E}_{\Theta_O}(\alpha_{in})$ $\{A_1, A_2, A_3, \Theta\}$	$\mathcal{E}_{\Theta_S}(\alpha_{in})$ $\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \Theta\}$
0	$\{0.75, 0.0, 0.0, 0.25\}$	$\{.180, .139, .110, .039, .073, .459\}$
30	$\{0.75, 0.0, 0.0, 0.25\}$	$\{.189, .128, .099, .033, .067, .484\}$
60	$\{0.75, 0.0, 0.0, 0.25\}$	$\{.175, .132, .091, .042, .070, .490\}$
90	$\{0.0, 0.75, 0.0, 0.25\}$	$\{.099, .173, .098, .092, .165, .373\}$
120	$\{0.0, 0.0, 0.0, 1.0\}$	$\{.178, .126, .106, .041, .066, .483\}$
150	$\{0.0, 0.0, 0.75, 0.25\}$	$\{.185, .133, .114, .059, .090, .419\}$

Table 5.12: Mass Assignment Models of the Evidence Generated by Optical and Sonar Sensors. Note that, $A_1 = \{\theta_1, \theta_2, \theta_3, \theta_6\}$, $A_2 = \{\theta_4, \theta_5, \theta_6\}$, $A_3 = \{\theta_1, \theta_2, \theta_3\}$.

Optical Source Alone: Here we study the use of the optical source alone. One can rotate around a target to obtain as much information as possible to characterize the object into one of the $\theta_i \in \Theta$. As one rotates around the object, one can update the existing belief about the object. This is illustrated under the column $\mathcal{E}_{\Theta}(\theta) = \mathcal{E}_{\Theta_O}(\theta - 1) \triangleleft \mathcal{E}_{\Theta_O}(\theta)$ in Table 5.13.

Initially, one can see that the evidence is more towards a target which is not spherical. As more views are considered, the support weighs more towards a cylindrical target. However, it is not possible to pin-point to one particular target. Also, notice the fact that there is significant amount of mass being assigned to Θ , representing the ambiguity in the available evidence. Thus, even after fusing 6 views together, one cannot make a crisp decision based on this evidence alone.

Acoustic Source Alone: Individual evidence corresponding to each α_{in} from the acoustic sensor is very uncertain. For instance, $\mathcal{E}_{\Theta_S}(60)$ leans more towards target type θ_1 , whereas $\mathcal{E}_{\Theta_S}(90)$ leans more towards target type θ_2 . However, the uncertainty in both of these BoEs is quite high as indicated by the high mass for complete ambiguity.

As before, let us explore the pattern of belief change as one updates its belief over different passes. One clear observation is that the mass assigned to total ambiguity decreases over the iterations. The support towards target type θ_1 increases. However, supports towards θ_2 and θ_3 are not negligible given the mass assigned to total ambiguity.

	$\mathcal{E}_{\Theta_O}(\alpha_{in}) \oplus \mathcal{E}_{\Theta_S}(\alpha_{in})$	$\mathcal{E}_{\Theta}(\alpha_{in} - 1) \triangleleft \mathcal{E}_{\Theta_O}(\alpha_{in})$	$\mathcal{E}_{\Theta}(\alpha_{in} - 1) \triangleleft \mathcal{E}_{\Theta_S}(\alpha_{in})$	$(\mathcal{E}_{\Theta}(\alpha_{in} - 1) \triangleleft \mathcal{E}_{\Theta_O}(\alpha_{in})) \dots \triangleleft \mathcal{E}_{\Theta_S}(\alpha_{in})$
α_{in}	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, A_1, A_2, A_3, \Theta\}$	$\{A_1, A_2, A_3, \Theta\}$	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \Theta\}$	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, A_1, A_2, A_3, \Theta\}$
0	$\{.196, .152, .120, .011, .020, .376, .000, .000, \mathbf{.501}\}$	$\{\mathbf{.750}, .000, .000, \mathbf{.250}\}$	$\{\mathbf{.179}, .139, .110, .039, .073, \mathbf{.460}\}$	$\{.131, .102, .080, .029, .054, \mathbf{.375}, .000, .000, \mathbf{.229}\}$
30	$\{\mathbf{.204}, .139, .107, .009, .018, \mathbf{.392}, .000, .000, .131\}$	$\{\mathbf{.844}, .000, .000, \mathbf{.156}\}$	$\{\mathbf{.230}, .165, .128, .044, .086, \mathbf{.347}\}$	$\{\mathbf{.181}, .129, .100, .029, .058, \mathbf{.358}, .000, .000, .145\}$
60	$\{\mathbf{.191}, .144, .099, .011, .019, \mathbf{.402}, .000, .000, .134\}$	$\{\mathbf{.891}, .000, .000, \mathbf{.109}\}$	$\{\mathbf{.246}, .181, .132, .053, .095, \mathbf{.293}\}$	$\{\mathbf{.198}, .146, .106, .034, .061, \mathbf{.352}, .000, .000, .103\}$
90	$\{.034, .060, .034, .128, \mathbf{.228}, .000, \mathbf{.387}, .000, .129\}$	$\{\mathbf{.445}, \mathbf{.469}, .000, .086\}$	$\{.190, \mathbf{.209}, .133, .090, .161, \mathbf{.217}\}$	$\{.140, .144, .093, .071, .127, \mathbf{.176}, \mathbf{.180}, .000, .069\}$
120	$\{\mathbf{.178}, .126, .105, .041, .066, .000, .000, .000, \mathbf{.484}\}$	$\{.223, \mathbf{.234}, .000, \mathbf{.543}\}$	$\{\mathbf{.227}, .198, .145, .075, .130, \mathbf{.225}\}$	$\{\mathbf{.177}, .148, .110, .060, .103, .088, .090, .000, \mathbf{.224}\}$
150	$\{\mathbf{.209}, .150, .128, .017, .250, .000, .000, \mathbf{.354}, .117\}$	$\{.111, .117, \mathbf{.469}, \mathbf{.303}\}$	$\{\mathbf{.245}, .193, .153, .080, .129, \mathbf{.200}\}$	$\{\mathbf{.196}, .151, .121, .056, .090, .044, .045, \mathbf{.163}, .134\}$

Table 5.13: Mass Models of Fused Evidence BoEs. Note that, $A_1 = \{\theta_1, \theta_2, \theta_3, \theta_6\}$, $A_2 = \{\theta_4, \theta_5, \theta_6\}$, $A_3 = \{\theta_1, \theta_2, \theta_3\}$. Highest and second highest masses in each BoE are marked in **bold** and underlined, respectively. These represent the propositions with the highest ‘support’.

Combining Both Optical and Acoustic Sources: The obvious fusion strategy is to combine evidence from each source for each α_{in} separately. The results obtained in this way are tabulated in the first column of Table 5.13. We have used DCR for combining \mathcal{E}_{Θ_S} and \mathcal{E}_{Θ_O} . One may also do the following: at each observation, update the belief by first fusing evidence from both optical and acoustic sources together and then using this combined evidence to update the existing belief. This is a more intuitive way of updating an existing belief. See

column $(\mathcal{E}_\Theta(\theta - 1) \triangleleft \mathcal{E}_{\Theta_o}(\theta)) \triangleleft \mathcal{E}_{\Theta_s}(\theta)$ in Table 5.13 for the BoEs obtained at each α_{in} . Observations are very similar to the previous cases in terms of the decreasing overall uncertainty. As more evidence is gathered, the support moves towards target type θ_1 and the overall uncertainty represented by $m(\Theta)$ decreases.

5.6 Chapter Summary

The CFE addressed the challenges **C₁**-**C₄** associated with soft/hard data fusion. The use of ported conditional notions allows the CFE to be represented in one unified setting for both identical and non-identical FoDs. CFE also possesses several intuitively appealing features which seem to indicate its suitability for scenarios that call for soft and hard evidence fusion. We have also shown how various fusion characteristics can be obtained via suitable parameter selections. Challenge **C₅**—computation overhead associated with conditionals and other DST computations are addressed in Chapters 7 and 8 under Part III: Computational Optimizations.

Consensus in a Fusion Environment

CONSENSUS is a fundamental issue arising in a fusion scenario. In this chapter, we study the consensus problem as a special case of finding *fixed-points* in asynchronous iterations for nonlinear *paracontracting operators* [83]. This allows for a unified analysis of consensus in arbitrary sensor networks with possible communication impairments. For instance, convergence of a *fully-connected* “perfect” network and an *ad-hoc* network with communication failures can both be analyzed using the same principles. Based on the CUE developed in Chapter 5, we present a complete theoretical development of a consensus protocol applicable to DST data fusion environments. Criteria for convergence analysis in several *network topologies* is provided thus addressing a wider class of distributed sensor networks.

Mathematical rigor, which requires somewhat complex notation, is essential for the development of a unified answer to convergence analysis covering a broad spectrum of networks. It is also very helpful in understanding the intricacies involved in such iterative infinite processes. We will utilize figures and remarks to better illustrate the salient points of the arguments and highlight the properties and features of the strategy being developed. Proofs of the results that are relevant to understanding the presentation appears within the text, while the rest of the proofs are relegated to the

appendix for clarity of the presentation. A credibility estimation technique, based on the proposed consensus notions, is developed in Chapter 7.

This chapter is organized as follows: Section 6.1 presents an overview of the basic notions of consensus, consensus protocols and network topologies on which the consensus is usually sought; Section 6.2 introduces the notions of asynchronous iterations, paracontractions and their convergence; and presents a reformulation of consensus protocols as asynchronous iterations; Section 6.3 presents the development of a new consensus protocol that is applicable to soft/hard data fusion environments; convergence criteria for several common network topologies are also derived; and finally, Section 6.4 contains the chapter summary. The proofs of the results that do not appear within the text are provided in Section A.3 of Appendix A.

6.1 Overview

The word consensus refers to a *general agreement* [26] among sources, e.g., a consensus of opinions among a jury pool. The notion of consensus is becoming increasingly popular among sensor related research/applications [84–90], where an agreement is sought among a group of agents¹³. Agents mutually exchange their states (which may represent an opinion or belief on a certain scenario), until they all converge. The way agents interact plays a key-role in reaching a consensus among their opinions.

6.1.1 Agent Interactions

Consider a set $\mathcal{N} = \{\mathcal{A}_1, \dots, \mathcal{A}_m\}$ of m interacting agents. A multi-agent system undergoes changes when an agent updates its state by interacting with other agents at discrete time instances $t_0 < t_1 < \dots < t_k < \dots$. Here, t_k is referred to as the *discrete event-based time* and k is referred to as the *discrete event-based time index*.

¹³The term ‘agent’ is often used in consensus literature to refer to sources, which can be either soft or hard. In the case of hard sensors, the change of an agent’s belief refers to the notion where a change of belief occurs in the fusion node managing the actual physical sensor, but not a change in the measurements/readings of the actual sensor.

Time t_k is discrete event-based time since it refers to a discrete event, not necessarily to an absolute time reference. Hence, without loss of generality, we can assume that only one agent updates its state at any event-based discrete time t_k and the sequence satisfies $t_0 < t_1 < \dots < t_k < \dots$. Indeed, in a situation where $n \leq m$ multiple agents $\{\mathcal{A}_{i_1}, \dots, \mathcal{A}_{i_n}\}$ update their states at the same time t_k , one can derive a corresponding discrete event-based time sequence $\dots < t_{k-1} < t_{k_1} \leq t_{k_2} \leq \dots \leq t_{k_n} < t_{k+1} \dots$ so that only agent \mathcal{A}_{i_j} updates its state at time t_{k_j} , for $j = 1, \dots, n$. See Figure 6.1. Hence, from now onwards, we assume that only a single agent updates the state at time t_k and the sequence satisfies $t_0 < t_1 < \dots < t_k < \dots$.

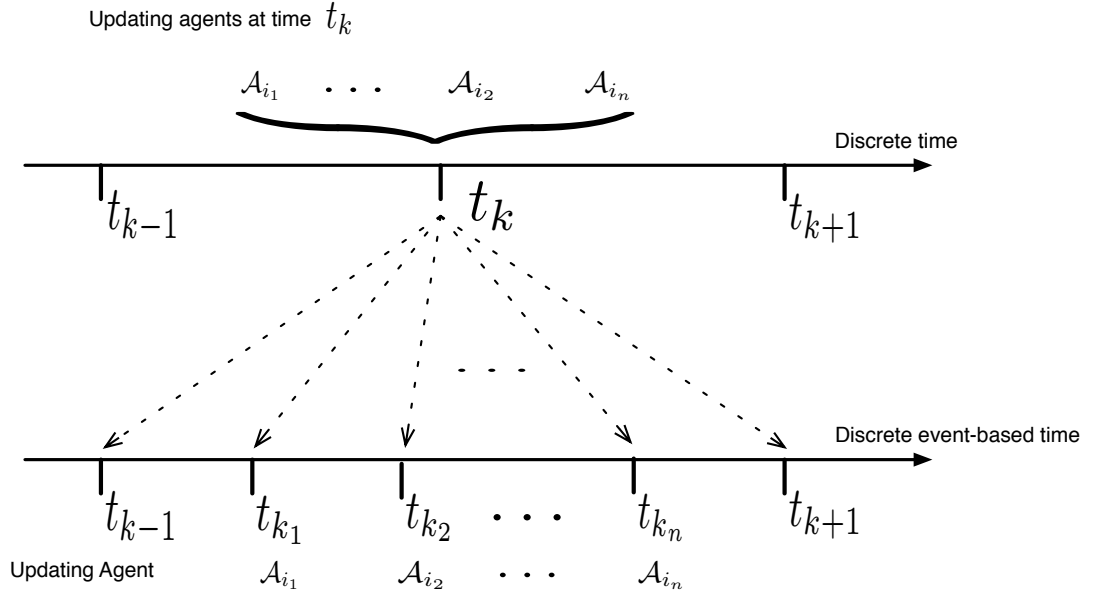


Figure 6.1: When the n agents $\{\mathcal{A}_{i_1}, \dots, \mathcal{A}_{i_n}\}$ update their states at the same time t_k , a discrete event-based time sequence $\dots < t_{k-1} < t_{k_1} \leq t_{k_2} \leq \dots \leq t_{k_n} < t_{k+1} \dots$ can be defined so that only agent \mathcal{A}_{i_j} updates its state at time t_{k_j} , for $j = 1, \dots, n$.

6.1.1.1 Agent Interaction Topology

Agent interaction topology refers to the structure of the spatial connectivity among agents at a discrete event-based time t_k . An agent, over the course of information exchange, may use different interaction topologies for state updates. Let us denote by $Q_{i,j}$, $j = 1, \dots, n_i$, the j^{th} interaction topology used by agent \mathcal{A}_i , for $i = 1, \dots, m$. Let us also use $\mathcal{Q} \equiv \{Q_{i,j} | j = 1, \dots, n_i; i = 1, \dots, m\}$ to denote the set of all interaction topologies used by the multi-agent system.

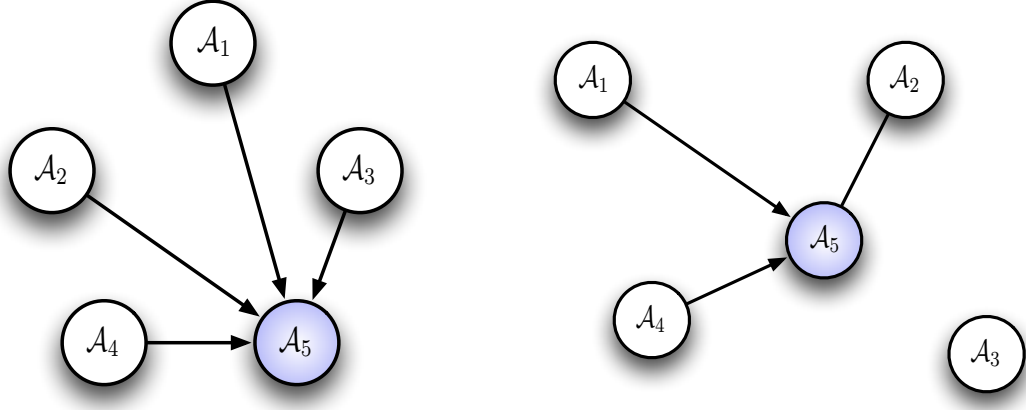
Since only a single agent updates its state at any given t_k , agent interaction topology at t_k can be modeled via a directed graph $G[k] \equiv (\mathcal{N}, E[k])$, where each edge $e_{ij} \in E[k]$ represents a unidirectional information exchange link from agent \mathcal{A}_i to agent \mathcal{A}_j , i.e., only \mathcal{A}_j can receive information from \mathcal{A}_i (see Figure 6.2). Let us use the sequence $\{J[k] \in \mathcal{Q} | k = 0, 1, \dots\}$ to identify the sequence of agent interactions, i.e., $J[k]$ identifies the interaction topology “active” at discrete event-based time index t_k , for $k = 0, 1, \dots$.

6.1.1.2 Fully-Connected Versus Partially-Connected Systems

An interaction topology is said to be *fully-connected* if the agent updating its state receives information from all the other agents (see Figure 6.2 (a)). A multi-agent system is fully-connected if all the interaction topologies (i.e., for all $Q \in \mathcal{Q}$) are fully-connected. If an agent updates its state without taking information from all the other agents, then the corresponding interaction topology is said to be *partially-connected* (see Figure 6.2 (b)). A multi-agent system is *partially-connected* if at least one interaction topology in use is partially-connected.

6.1.1.3 Static Versus Dynamic Systems

Multi-agent system can be either *static* or *dynamic*. If the interaction topologies of each agent \mathcal{A}_i is time invariant for all the agents $\mathcal{A}_1, \dots, \mathcal{A}_n$, then we refer to such



(a) $Q_{5,1}$: a *fully-connected* interaction topology (b) $Q_{5,2}$: a *partially-connected* interaction topology

Figure 6.2: Spatial connectivity among a set of agents $\mathcal{N} = \{A_1, \dots, A_5\}$. In (a), agent A_5 receives information from all other agents; hence, the interaction topology is fully-connected. In (b), agent A_5 does not receive information from agent A_3 ; hence, it is only partially-connected.

a system as a **static multi-agent system**. Hence, in a static multi-agent system, we must have $Q_{i,j} = Q_i, j = 1, \dots, n_i$ for all $i = 1, \dots, m$, where Q_i denotes the interaction topology of agent A_i for $i = 1, \dots, m$. If this does not hold true (i.e., at least one agent uses interaction topologies $Q_{i,j}, Q_{i,k}$ s.t. $Q_{i,j} \neq Q_{i,k}$), then we refer to such a system as a **dynamic multi-agent system** (see Figure 6.3).

6.1.2 Consensus

The consensus problem can be formally stated as:

Definition 22 (Consensus). [88] Consider a set $\mathcal{N} = \{A_1, \dots, A_m\}$ of agents embedded at each discrete event-based time $t_k, k = 0, 1, \dots$, in an agent interaction topology modeled via a directed graph $G[k] = (\mathcal{N}, E[k])$, where each edge $e_{ij} \in E[k]$ represents a unidirectional information exchange link from agent A_i to agent A_j . Each agent $A_i \in \mathcal{N}$ starts with the initial state $\mathbf{x}_i[0] \in \mathbb{D} \subseteq \mathbb{R}^n$ and repeatedly updates its state

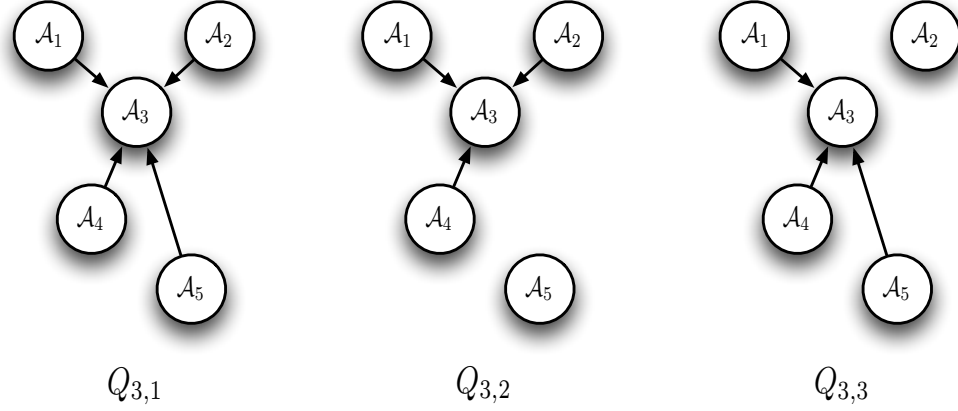


Figure 6.3: A dynamic partially-connected 5 agent network which shows three different interaction topologies of agent \mathcal{A}_3 .

by exchanging information according to the communication network specified by $G[k]$.

We say that a consensus is reached among the agents in \mathcal{N} , if $\|\mathbf{x}_i[k] - \mathbf{x}_j[k]\| \rightarrow 0$ as $k \rightarrow \infty$ for all $\mathcal{A}_i, \mathcal{A}_j \in \mathcal{N}$, for some norm $\|\cdot\| : \mathbb{D} \mapsto \mathbb{R}$. \square

6.1.3 Consensus Protocols

The consensus generating mechanisms are referred to as *consensus control strategies*, *consensus control protocols* or simply *consensus protocols*.

6.1.3.1 Linear Versus Non-Linear Consensus Protocols

In linear consensus protocols, the agents update their states as a *convex sum* of the states of agents. For instance,

$$\mathbf{x}_i[k+1] = \sum_{j=1}^m f_{ij}[k] \mathbf{x}_j[k], \quad (6.1)$$

where $f_{ij}[k] \geq 0$ and $\sum_{j=1}^m f_{ij}[k] = 1$ for all i, j , specifies a general linear consensus protocol. Here, whenever $f_{ij}[k] > 0$, agent \mathcal{A}_j communicates its current state to agent

\mathcal{A}_i , weighted by $f_{ij}[k]$. However, in the case of a non-linear protocol, the state update between agents is non-linear. For instance,

$$\mathbf{x}_i[k+1] = F^{ij}(\mathbf{x}_1[k], \dots, \mathbf{x}_m[k]) \quad (6.2)$$

specifies a general non-linear consensus protocol, where the functions $F^{ij}(\cdot, \dots, \cdot)$, $j = 1, \dots, n_i$ for $i = 1, \dots, m$ specify the non-linear interaction among agents¹⁴. Clearly, linear consensus protocols are a special case of non-linear consensus protocols.

6.1.3.2 Synchronous Versus Asynchronous Consensus Protocols

These consensus protocols can be either *synchronous* or *asynchronous*. The consensus protocols in (6.1) and (6.2) are synchronous, in the sense that all the agents update their states using the latest state values of the other agents (i.e., there is no lag in communication paths). However, due to various imperfections in communications, such as link failures and synchronization errors, etc., synchronous updates may not be possible in real-life situations, thus generating asynchronous consensus protocols. Asynchronous consensus protocols can be formally defined as follows.

Definition 23 (Non-Linear Asynchronous Consensus Protocol). [88] *Let $t_0 < t_1 < \dots < t_k < \dots$ denote the discrete event-based time instances that the state of a multi-agent system undergoes changes. Let $\mathbf{x}_i[k]$ denote the state of agent \mathcal{A}_i at time t_k for $k = 0, 1, \dots$. Let $\mathcal{Q} = \{Q_{i,j} \mid j = 1, \dots, n_i, i = 1, \dots, m\}$ denote the set of all interaction topologies observed during the consensus process. Here, $Q_{i,j}$ denotes the j^{th} interaction topology used by agent \mathcal{A}_i , for $j = 1, \dots, n_i$ and $i = 1, \dots, m$. Let $J[k] \in \mathcal{Q}$ identify the agent interactions at time t_k , for $k = 0, 1, \dots$. Then, an*

¹⁴Here, each agent \mathcal{A}_i updates its states via one of the m_i functions available, for $i = 1, \dots, m$.

asynchronous consensus protocol is given by

$$\mathbf{x}_i[k+1] = \begin{cases} F^{i_j}(\mathbf{x}_1(s^1[k]), \dots, \mathbf{x}_m(s^m[k])), & \text{for } J[k] = Q_{i,j}; \\ \mathbf{x}_i[k], & \text{otherwise,} \end{cases} \quad (6.3)$$

where F^{i_j} specifies the non-linear interaction topology corresponding to $Q_{i,j}$ and $s^\ell[k] \in \mathbb{N}_0 \equiv \{0, 1, \dots\}$ s.t. $s^\ell[k] \leq k$ for $\ell = 1, \dots, m$, $k = 0, 1, \dots$. The quantities $k - s^\ell[k]$ and $J[k]$ are referred to as **iteration delay** and **updating sets**, respectively. \square

Following remarks are noteworthy.

- Definition 23 describes the most general consensus protocol. It is applicable to any multi-agent system that is partially or fully connected, static or dynamic, and having synchronous or asynchronous communications.
- At a given time instance t_k , the agent interaction is fully characterized by $J[k]$. For example, if $J[k] = Q_{3,8}$ then the agent \mathcal{A}_3 updates its states via $F^{38}(\cdot, \dots, \cdot)$, while all the other agents will retain the previous states (i.e., $\mathbf{x}_i[k+1] = \mathbf{x}_i[k]$, for $i \neq 3$). Hence, the sequence $\{J[k] \mid k = 0, 1, \dots\}$ identifies the sequence of agent interactions at discrete event-based time instances $t_0 < t_1 < \dots$.
- Note that $F^{i_j}(\cdot, \dots, \cdot)$ could be any non-linear or linear function. As we will introduce in the following section, the class of functions that is referred to as *paracontracting operators* is of special importance in the study of consensus.
- Also, notice that synchronous consensus protocols are a special case of asynchronous consensus protocols (i.e., when $s^\ell[k] = k$, for all ℓ, k). Thus, non-linear asynchronous consensus protocols subsume all other consensus protocols.

6.2 Paracontractions View of Consensus Protocols

Consensus analysis in multi-agent systems can be formulated as a special case of finding common *fixed points* of a (finite) pool of *paracontracting* multiple-point operators [88]. Convergence of these schemes can then be established if the iterations are bounded by certain *coupling* conditions. Let us proceed by first introducing the basic notions of the theory of paracontractions.

6.2.1 Paracontracting Operators

For the purpose at hand, we use \mathbb{D} to denote the ‘domain of interest.’ In consensus analysis, one can take \mathbb{D} to be the domain of agent states, which can be a real vector, probability mass function or any other valid and appropriate representation. Consensus protocols developed in this chapter consider agent states modeled as DS BoEs. So, for this particular case, \mathbb{D} is the domain of all valid DS BoEs.

Definition 24 (Fixed point). *Let \mathbb{D} be the domain of interest. A vector $\xi \in \mathbb{D}$ is referred to as a **fixed point** of an operator $F : \mathbb{D}^m \mapsto \mathbb{D}$ iff $F(\xi, \dots, \xi) = \xi$, where $m \in \mathbb{N}$. Further, the set of all fixed points of operator F is denoted by $\text{fix}(F)$, where $\text{fix}(F) \equiv \{\xi \in \mathbb{D} \mid F(\xi, \dots, \xi) = \xi\}$.*

*A vector $\zeta \in \mathbb{D}$ is a **common fixed point** of a pool of operators \mathcal{F} , if ζ is a fixed point common to all operators $F \in \mathcal{F}$, viz., $\zeta \in \text{fix}(F)$, $\forall F \in \mathcal{F}$.* □

We now introduce

Definition 25 (Paracontracting Operators). *[91] An operator $F : \mathbb{D}^m \mapsto \mathbb{D}$ is **paracontractive** on \mathbb{D} with respect to a given vector norm $\|\cdot\|$, if*

$$\|F(\mathbf{X}) - \xi\| < \max_j \|\mathbf{x}_j - \xi\|, \quad (6.4)$$

for all $\mathbf{X} \equiv [\mathbf{x}_1, \dots, \mathbf{x}_m] \in \mathbb{D}^m$ and any $\xi \in \text{fix}(F)$, unless $\mathbf{X} \in \text{fix}(F)$. □

Definition 26 (Paracontracting Pool of Operators). [91] Let $\mathbb{I} \subset \mathbb{N} \equiv \{1, 2, \dots\}$ be a set of indices, $m \in \mathbb{N}$ a fixed number and $\mathcal{F} = \{F^i, i \in \mathbb{I} \mid F^i : \mathbb{D}^{m_i} \mapsto \mathbb{D}\}$ be a pool of operators, where $m_i \in \mathbb{N}$ s.t. $m_i \leq m, \forall i \in \mathbb{I}$, and \mathbb{D} is closed. Then, if for all $i \in \mathbb{I}$, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_{m_i}] \in \mathbb{D}^{m_i}$ and a vector norm $\|\cdot\|$, F^i is continuous on \mathbb{D}^{m_i} , then \mathcal{F} is said to be **paracontracting** on \mathbb{D} , if for any $\xi \in \text{fix}(F^i)$,

$$\|F^i(\mathbf{X}) - \xi\| < \max_j \|\mathbf{x}_j - \xi\|, \quad (6.5)$$

unless $\mathbf{X} \in \text{fix}(F^i)$. □

Example 11. Consider a pool of operators $\mathcal{F} = \{F^i \mid i = 1, \dots, 6\}$, where the operators are defined as $F^1 : \mathbb{R}^2 \mapsto \mathbb{R}$, $F^2 : \mathbb{R}^2 \mapsto \mathbb{R}$, $F^3 : \mathbb{R}^4 \mapsto \mathbb{R}$, $F^4 : \mathbb{R}^3 \mapsto \mathbb{R}$, $F^5 : \mathbb{R}^2 \mapsto \mathbb{R}$ and $F^6 : \mathbb{R}^3 \mapsto \mathbb{R}$. The set $\mathbb{I} = \{1, \dots, 6\}$ is an index set for all the operators.

So, if all the operators are continuous and paracontracting on \mathbb{R} w.r.t. Definition 25, then, according to Definition 26, the whole pool of operators \mathcal{F} is paracontractive on \mathbb{R} . ■

The above defines the paracontraction property for a pool of operators. As we will explain in the following sections, a pool of operators is used to update a multi-agent system where each operator is a paracontraction on the domain of interest¹⁵.

6.2.2 Non-Linear Asynchronous Iterations

An asynchronous iteration can be formally defined as follows.

¹⁵The operators in Example 11 were selected with this idea in mind. Notice the connection of F^i to the node connectivity of Agent \mathcal{A}_i of Figure 6.2 (b).

Definition 27 (Asynchronous Iteration). [83] Let \mathbb{I} be a set of indices, $m \in \mathbb{N}$ a fixed number and $\mathcal{F} = \{F^i, i \in \mathbb{I} \mid F^i : \mathbb{D}^{m_i} \mapsto \mathbb{D}, m_i \in \mathbb{N} \text{ s.t. } m_i \leq m\}$ be a pool of operators, where \mathbb{D} is closed. Let $\mathcal{X}_0 = \{\mathbf{x}[-\ell] \in \mathbb{D} \mid \ell = 1, \dots, M\}$ be a given set of vectors, where M is the number of initial conditions.

Let \mathcal{S} denote the sequence of m_i -tuples from $\mathbb{N}_0 \cup \{-1, \dots, -M\}$, where $\mathcal{S} \equiv \left\{ s^1[k], \dots, s^{m_{I[k]}}[k] \mid I[k] \in \mathbb{I}, s^\ell[k] \in \mathbb{N}_0 \cup \{-1, \dots, -M\} \text{ s.t. } s^\ell[k] \leq k, k = 0, 1, \dots \right\}$. Then, for sequences $\mathcal{I} \equiv \{I[k] \in \mathbb{I} \mid k = 0, 1, \dots\}$ and \mathcal{S} , the sequence given by

$$\mathbf{x}[k+1] = F^{I[k]} \left(\mathbf{x}[s^1[k]], \dots, \mathbf{x}[s^{m_{I[k]}}[k]] \right), \quad (6.6)$$

is referred to as an **asynchronous iteration** and denoted by $(\mathcal{F}, \mathcal{X}_0, \mathcal{I}, \mathcal{S})$. \square

Example 12. Let us consider the pool of operator \mathcal{F} in Example 11. Also, consider $\mathcal{X}_0 = \{\mathbf{x}[-\ell] \mid \ell = 1, \dots, 6\}$ (so that we have $M = 6$ initial conditions). Here, the sequence $\mathcal{I} \equiv \{I[k] \in \mathbb{I} \mid k = 0, 1, \dots\}$ consists of elements of $\mathbb{I} = \{1, \dots, 6\}$. Consider the sequence $\mathcal{I} = \{1, 2, 4, 5, 3, 6, 1, \dots\}$. Hence, the iteration is done in the following order:

$$\begin{aligned} \mathbf{x}[1] &= F^1(\mathbf{x}[s^1[0]], \mathbf{x}[s^2[0]]), & k &= 0; \\ \mathbf{x}[2] &= F^2(\mathbf{x}[s^1[1]], \mathbf{x}[s^2[1]]), & k &= 1; \\ \mathbf{x}[3] &= F^4(\mathbf{x}[s^1[2]], \mathbf{x}[s^2[2]], \mathbf{x}[s^3[2]]), & k &= 2; \\ \mathbf{x}[4] &= F^5(\mathbf{x}[s^1[3]], \mathbf{x}[s^2[3]]), & k &= 3; \\ \mathbf{x}[5] &= F^3(\mathbf{x}[s^1[4]], \mathbf{x}[s^2[4]], \mathbf{x}[s^3[4]], \mathbf{x}[s^4[4]]), & k &= 4; \\ \mathbf{x}[6] &= F^6(\mathbf{x}[s^1[5]], \mathbf{x}[s^2[5]], \mathbf{x}[s^3[5]]), & k &= 5; \\ \mathbf{x}[7] &= F^1(\mathbf{x}[s^1[6]], \mathbf{x}[s^2[6]]), & k &= 6; \end{aligned}$$

$$\vdots$$

$$\mathbf{x}[k+1] = F^{I[k]}(\mathbf{x}[s^1[k]], \dots, \mathbf{x}[s^{m_{I[k]}}[k]]), \quad k.$$

The sequence \mathcal{S} , in a loose sense, defines the “time lag” in updating. To illustrate, consider the sequence $\mathcal{S} \equiv \{(s^1[0] = -2, s^2[0] = -1), (s^1[1] = -1, s^2[1] = 0), (s^1[2] = -2, s^2[2] = 0, s^3[2] = -5), (s^1[3] = -4, s^2[3] = 2), \dots\}$. Then, the first 4 iterations are given by

$$\mathbf{x}[1] = F^1(\mathbf{x}[-2], \mathbf{x}[-1]), \quad k = 0;$$

$$\mathbf{x}[2] = F^2(\mathbf{x}[-1], \mathbf{x}[0]), \quad k = 1;$$

$$\mathbf{x}[3] = F^4(\mathbf{x}[-2], \mathbf{x}[0], \mathbf{x}[-5]), \quad k = 2;$$

$$\mathbf{x}[4] = F^5(\mathbf{x}[-4], \mathbf{x}[2]), \quad k = 3.$$

Notice that $s^\ell(k) \leq k$ for all k . ■

6.2.3 Confluence and Regulatory Assumptions

Convergence of an asynchronous iteration scheme is dependent on the properties of the operators and the *coupling* among the agents. Let us study the restrictions on coupling that must be imposed to yield convergence in the sense of Theorem 14 (see Section 6.2.4).

Definition 28. [91] Let $(\mathcal{F}, \mathcal{X}_0, \mathcal{I}, \mathcal{S})$ be an asynchronous iteration. Then,

(i) \mathcal{S} is **regulated** if

$$s \equiv \max_{k, \ell} (k - s^\ell[k]) \text{ exists.}$$

(ii) \mathcal{I} is an **index-regulated sequence** if, for all $i \in \mathbb{I}$, $\exists c_i \in \mathbb{N}_0$, s.t. for all $k \in \mathbb{N}_0$,

$$i \in \{I[k]\} \cup \{I[k+1]\} \cup \dots \cup \{I[k+c_i]\}.$$

(iii) \mathcal{I} is **regulated** if $\exists c \in \mathbb{N}_0$, s.t. for all $k \in \mathbb{N}_0$,

$$\{I[k]\} \cup \{I[k+1]\} \cup \dots \cup \{I[k+c]\} = \mathbb{I}.$$

Let us explain these notions via an example.

Example 13. Consider Example 12 again.

To check if the sequence \mathcal{S} is regulated, we need to check all of $s^\ell[k]$ to see if (i) is satisfied. Here, we have $\max_{k,\ell}(k - s^\ell[k]) = 7$, for $k = 0, \dots, 7$. Similarly, if $\max_{k,\ell}(k - s^\ell[k])$ is finite for all k, ℓ , then the sequence \mathcal{S} is regulated.

To check if the sequence \mathcal{I} is index-regulated, from a starting position k , we should be able to locate finite segments of the sequence \mathcal{I} that contains i , for all $i \in \mathbb{I}$. If this is satisfied for all k , then the \mathcal{I} is index-regulated. Here, we have $1 \in \{I[0]\}$, $2 \in \{I[0] \cup I[1]\}$, $3 \in \{I[0] \cup I[1] \cup I[2] \cup I[3] \cup I[4]\}$, $4 \in \{I[0] \cup I[1] \cup I[2]\}$, $5 \in \{I[0] \cup I[1] \cup I[2] \cup I[3]\}$ and $6 \in \{I[0] \cup I[1] \cup I[2] \cup I[3] \cup I[4] \cup I[5]\}$. So, \mathcal{I} satisfies (ii) for $k = 0$. If \mathcal{I} continues to satisfy (ii) for all $k \geq 0$, then it is index-regulated.

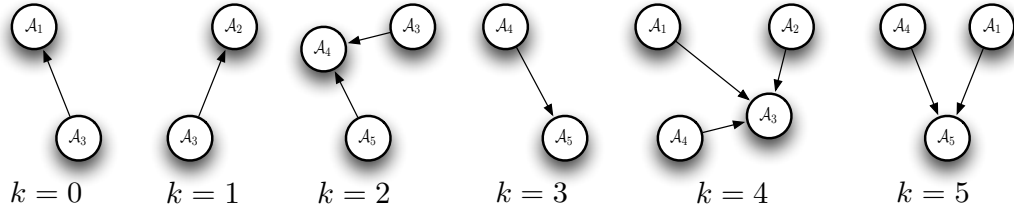
To check if the sequence \mathcal{I} is regulated, from an arbitrary starting position $k \geq 0$, we must find a finite segment of the sequence that contains \mathbb{I} . Here, we have $\{I[0] \cup I[1] \cup I[2] \cup I[3] \cup I[4] \cup I[5]\} = \mathbb{I}$. If \mathcal{I} continues to satisfy (iii) for all $k \geq 0$, then \mathcal{I} is regulated. ■

The set of edges $E[k]$ in graph $G[k] \equiv (\mathcal{N}, E[k])$ defines the *spatial* coupling among agents at a given discrete-event based index k . However, in an asynchronous iteration process, the coupling among agents at various time instances must be constrained to guarantee convergence. This is perhaps best explained via the notion of an *iteration graph* [88].

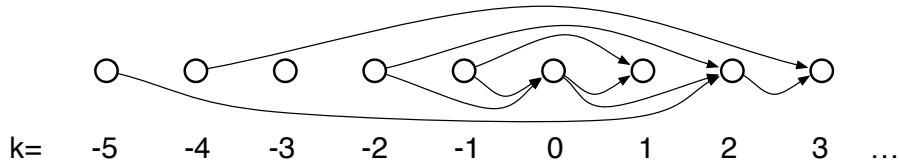
Definition 29 (Iteration Graph). *The iteration graph of the asynchronous iteration $(\mathcal{F}, \mathcal{X}_0, \mathcal{I}, \mathcal{S})$ is the digraph $(\mathcal{V}, E_{\mathcal{S}})$, where the set of vertices and edges are given by $\mathcal{V} \equiv \mathbb{N}_0 \cup \{-1, \dots, -M\}$ and $E_{\mathcal{S}} \equiv \{(k, k_0) \mid \exists 1 \leq \ell \leq m_{I(k_0-1)} \text{ s.t. } s^\ell(k_0 - 1) = k\}$, respectively. \square*

Let us illustrate these notions via an example.

Example 14. *Consider a multi-agent system consisting of 5 agents, \mathcal{A}_i , $i = 1, \dots, 5$, whose interactions are described by an asynchronous iteration $(\mathcal{F}, \mathcal{X}_0, \mathcal{I}, \mathcal{S})$, where \mathcal{F} as given in Example 11, and the initial conditions \mathcal{X}_0 and the sequences \mathcal{I} and \mathcal{S} as given in Example 12. Assume that agent \mathcal{A}_i updates the vector \mathbf{x} using operator F^i for $i = 1, \dots, 4$ and agent \mathcal{A}_5 uses operators F^5 and F^6 . Hence, at a given k , the sequence \mathcal{I} determines, which agent performs the update on $\mathbf{x}[k]$. This (spatial) interaction is shown in the interaction graph in Figure 6.4 (a).*



(a) Interaction Graph



(b) Iteration Graph

Figure 6.4: Interaction and iteration graphs for Example 14.

However, note that the iteration delays in the updating scheme cannot be identified

via the interaction graph. As we illustrate later, adequate coupling in terms of iteration is important for convergence in asynchronous networks. Figure 6.4 shows the iteration graph of this scheme; here, each node represents an iteration index. Nodes -5 through -1 represents the initial conditions \mathcal{X}_0 . A node that represents the iteration index k has edges coming to it from the iterations of which the state vectors are used by the current iteration. For example, node $k = 1$ has edges coming from $k = 0$ and $k = -1$, since the $\mathbf{x}[1 + 1]$ uses the states $\mathbf{x}[0]$ and $\mathbf{x}[-1]$. ■

The coupling conditions are imposed to guarantee the existence of convergence. In addition, the *confluent conditions* are imposed to make sure that all the operators in \mathcal{F} are involved in the iteration process. An iteration that satisfies confluent conditions can be formally defined as:

Definition 30 (Confluent Iteration). *An asynchronous iteration $(\mathcal{F}, \mathcal{X}_0, \mathcal{I}, \mathcal{S})$ is **confluent** if there are numbers $n_0, b \in \mathbb{N}$ and a sequence $\{b_k \in \mathbb{N} \mid k = n_0, n_0 + 1, \dots \text{ s.t. } k \geq n_0\}$ s.t. the following is true:*

- (i) *for every vertex $k_0 \geq k$, there is a path from b_k to k_0 in $(\mathcal{V}, \mathcal{E})$;*
- (ii) *$k - b_k \leq b$;*
- (iii) *\mathcal{S} is regulated;*
- (iv) *for every $i \in \mathcal{I}$, there is a $c_i \in \mathbb{N}$, so that for all $k \geq n_0$, there is a vertex $w_k^i \in \mathcal{V}$, which is a successor of b_k and a predecessor of b_{k+c_i} , for which $I(w_k^i - 1) = i$. □*

6.2.4 Convergence Criteria

The convergence of non-linear asynchronous iterations are governed by

Theorem 14. *Let \mathcal{F} be a paracontracting pool on $\mathbb{D} \subset \mathbb{R}^n$ and assume that \mathcal{F} has common fixed points, viz., $\text{fix}(\mathcal{F}) \neq \emptyset$. Then, any confluent asynchronous iteration $(\mathcal{F}, \mathcal{X}_0, \mathcal{I}, \mathcal{S})$ converges to some fixed point $\xi \in \text{fix}(\mathcal{F})$. \square*

Hence, an asynchronous iteration involving a pool of multiple-point operators converges if (i) the operators are paracontracting, (ii) pool has common fixed points and (iii) the iterations among operators are confluent. Next, we study how these results can be applied for analysis of asynchronous consensus protocols.

6.2.5 Non-Linear Asynchronous Consensus Protocols

If the consensus protocols take the form of an asynchronous iteration, we can make use of the above convergence results to develop consensus protocol by using paracontracting operators and posing necessary coupling conditions for confluent iterations. However, the relationship between consensus protocols and asynchronous iterations is not obvious from the definitions¹⁶. We now reformulate the non-linear asynchronous consensus protocol in Definition 23 as an asynchronous iteration.

6.2.5.1 Consensus Protocols as Asynchronous Iterations

In an asynchronous iteration, at index k , a vector $\mathbf{x}[k]$ is updated via an operator F^i , $i \in \mathbb{I}$ specified by $I[k]$. However, in a consensus protocol, a state vector of an agent specified by $J(k)$ is updated at time t_k , i.e. the state vector $\mathbf{x}_{i'}[k]$ corresponding to the active agent interaction topology $J[k] = Q_{i',j}$ at discrete-event based time t_k . The idea behind this transformation is to renumber the sequences in such a manner that updating vector $\mathbf{x}[k]$ at k via F^i is equivalent to updating $\mathbf{x}_{i'}[k']$ via F^i at some index k' , where $t_k = t'_{k'}$. Note that having $k \neq k'$ is irrelevant as long as $t_k = t_{k'}$.

¹⁶Notice that in (6.6), an $\mathbf{x} \in \mathbb{D} \subseteq \mathbb{R}^n$ is updated at every k , whereas in (6.3), state of agent \mathcal{A}_i which is given by $\mathbf{x}_i \in \mathbb{D} \subseteq \mathbb{R}^n$ is updated, for some $i \in \{1, \dots, m\}$. Hence, asynchronous iterations look at an update of a whole vector $\mathbf{x} \in \mathbb{D}$ at each discrete-event index k , compared to independent updates of individual elements in the case of consensus protocols.

$I[k]$ identifies the “active” operator F^i , $i \in \mathbb{I}$ at k ; also, each F^i , for all $i \in \mathbb{I}$ identifies the “active” interaction topology $Q_{i',j}$, which in turn identifies the agent $\mathcal{A}_{i'}$ that is updating its state. Hence, for ease and clarity of presentation, we assume that $I[k]$ directly identifies the active agent. So, here, $\mathbf{x}_{I[k]}[k] = \mathbf{x}_{i'}[k]$. We can now formulate the consensus protocol given in (6.3) to the form of an asynchronous iteration given in (6.6) as follows.

Step⁽¹⁾: Re-number $s^\ell[k]$, $k = 0, 1, \dots$ in such a manner that all the components $x_\ell[s^\ell[k]]$ in (6.3) are updated at time $s^\ell[k]$, i.e.,

$$I(s^i[k] - 1) = i, \quad \forall k \in \mathbb{N}, \quad i \in \{1, \dots, m\} \quad \text{with } s^i(k) \geq 1. \quad (6.7)$$

Step⁽²⁾: Assume all initial vectors are multiples of $\mathbf{1}$ and set

$$x[-k] := x_k[0]\mathbf{1}, \quad \forall k = 1, \dots, m; \quad (6.8)$$

and renumber the elements of sequences of $s^\ell[k]$, $k = 0, 1, \dots$ and $\ell = 1, \dots, m$ for which, $s^\ell[k] = 0$ in the same way.

Step⁽³⁾: Generate asynchronous iteration $(\mathcal{F}, \mathcal{Y}_\mathcal{O}, \mathcal{I}, \mathcal{S})$ with

$$y[k+1] := F^{I(k)}\left(y[\tilde{s}^1[k]], \dots, y[\tilde{s}^{m_{I[k]}}[k]]\right), \quad k = 0, 1, \dots \quad (6.9)$$

where $\mathcal{F} = \{F^{I[k]} \mid k = 0, 1, \dots\}$ as in (6.3), $\mathcal{I} = \{I[k], k = 0, 1, \dots\}$ and $\mathcal{S} = \{\tilde{s}^i[k] \mid k = 0, 1, \dots; i = 1, \dots, m_{I[k]}\}$ is given by

$$\tilde{s}^i[k] := s^{m_{I[k]}(i)}[k], \quad \forall k \in \mathbb{N}_0, \quad i = 1, \dots, m_{I[k]}, \quad (6.10)$$

and $\mathcal{Y}_\mathcal{O}$ by

$$y[-\ell] := x[-\ell], \quad \ell = 1, \dots, m. \quad (6.11)$$

\implies The asynchronous iteration in (6.9) and (6.10) is equivalent to the consensus protocol (6.3) in the sense that

$$y[k+1] = x_{I(k)}[k+1], \quad \forall k \in \mathbb{N}_0. \quad (6.12)$$

With this in place, we now establish the conditions for analysis of confluence in order to use the Theorem 14.

6.2.5.2 Verification of Confluence

Conditions for confluent iterations can be best described via the iteration graphs. Let us proceed by introducing some basic notions on graph composition [87].

Definition 31. Let \mathcal{G} be the set of all directed graphs with vertex set \mathcal{N} . Let $G_i \in \mathcal{G}$ be two arbitrary graphs defined as $G_i \equiv (N_i, E_i)$, where N_i and E_i are the sets of vertices and edges of G_i for $i = 1, 2$, respectively. The composition $G_2 \circ G_1$ is the graph $G_2 \circ G_1 = (N_{1 \circ 2}, E_{1 \circ 2})$ s.t. if $(u, w) \in E_{1 \circ 2}$ then $(u, v) \in E_1$ and $(v, w) \in E_2$.

A vertex $i \in N_1$ is a root of graph G_1 , if the edge $(i, j) \in E_1$ for all the vertices $j \in N_1 \setminus i$. A rooted graph has at least one root. A finite sequence of directed graphs $G_{p_1}, G_{p_2}, \dots, G_{p_k}$ from \mathcal{G} is jointly rooted, if the composition $G_{p_k} \circ G_{p_{k-1}} \circ \dots \circ G_{p_1}$ is a rooted graph. An infinite sequence of graphs G_{p_1}, G_{p_2}, \dots in \mathcal{G} is a repeatedly jointly rooted, if there is a positive integer r for which each finite sequence $G_{p_{r(k-1)+1}}, \dots, G_{p_{rk}}$, $k \geq 1$ is jointly rooted. \square

Example 15. Consider the set of graphs $G(i)$, $i = 1, \dots, 4$ as given in Figure 6.5. The graph $G(i)$ is rooted, since there is a path from node \mathcal{A}_i to all the other nodes, for $i = 1, \dots, 4$.

Now, the composition $G(4) \circ G(3) \circ G(2) \circ G(1)$ generates the edge set $\{(\mathcal{A}_1, \mathcal{A}_1), (\mathcal{A}_1, \mathcal{A}_2), (\mathcal{A}_1, \mathcal{A}_3)\}$, hence $G(4) \circ G(3) \circ G(2) \circ G(1)$ is a rooted graph, therefore the sequence $G(1), G(2), G(3), G(4)$ is jointly rooted. Also, as it turns out, the composition $G(4) \circ G(3) \circ G(2) \circ G(1) \equiv G(1)$. Hence, for any sequence containing this pat-

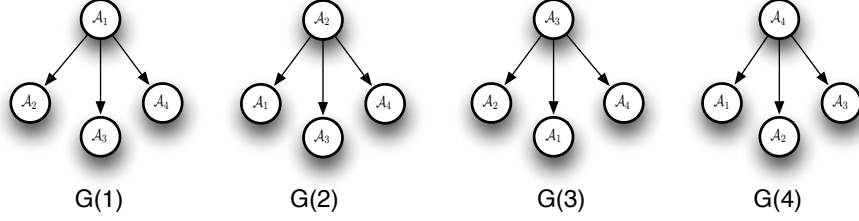


Figure 6.5: Example 15: An illustration of graph composition

tern will have a subsequence that is jointly rooted. Therefore, the infinite sequence, $\dots, G(4), G(1), G(2), G(3), G(4), G(1), \dots$ is repeatedly jointly rooted. ■

Graph composition is fundamental to understanding the coupling among operators, especially in an asynchronous situation. See Figure xxx for example of rooted and repeatedly jointly rooted graphs. A result that establishes the conditions for an asynchronous iteration to be confluent is given by:

Lemma 15. [88] Let $(\mathcal{F}, \mathcal{V}_O, \mathcal{I}, \mathcal{S})$ be the asynchronous iteration in (8). Consider any trajectory of $(\mathcal{F}, \mathcal{V}_O, \mathcal{I}, \mathcal{S})$ along the sequence of interaction graphs $G(0), G(1), \dots$. If the sequence $G(0), G(1), \dots$ is repeatedly jointly rooted; and satisfies the assumptions

- i.) Agent \mathcal{A}_{i_0} always uses its own latest state to update its current state. That is, $s^{i_0}(k) = \max\{k_0 \leq k \mid I_{k_0-1} = i_0\}$ for all $k > \min\{k_0 \in \mathbb{N}_0 \mid I_{k_0} = i_0\}$ with $I_k = i_0$;
- ii.) $\mathcal{I} = I_k, k = 0, 1, \dots$ is regulated;
- iii.) $k - s^\ell(k) \leq s, \forall k \in \mathbb{N}_0, \ell = 1, \dots, m$, for some $s \in \mathbb{N}_0$;

then, the asynchronous iteration $(\mathcal{F}, \mathcal{V}_O, \mathcal{I}, \mathcal{S})$ is confluent. □

6.2.5.3 Verification of Convergence

Convergence of an asynchronous consensus protocol can now be established by verifying, (i) the pool of operators to be paracontractive on the domain of interest; and (ii) the coupling conditions of an equivalent asynchronous iterations. With these developments in place, we are now in a position to develop a DST consensus protocol that is applicable for soft/hard data fusion environments.

6.3 Consensus in DST Fusion Environments

The notion of consensus has been used across several disciplines over several decades. The seminal work [92] of Lehrer and Wagner—professors of philosophy and mathematics, respectively—laid the foundation for mathematical modeling of consensus and its applications.

One of the basic rules of the consensus notion established by the work in [92], which became popular as *Lehrer-Wagner Model of rational consensus*, is that the consensus has to be established as a weighted average of opinions of agents. Most of the consensus protocols and techniques that are out there satisfy this condition, since they are weighted averages of rather simple numerical models or at most probability mass functions, and only provide very limited functionality. However, soft/hard data fusion environments call for rather sophisticated mechanisms that allow for one to incorporate elaborate features. For instance, in soft/hard fusion environments one often has access to highly reliable estimates of the Ground Truth (GT); these estimates are often vague, but highly reliable, e.g., prevailing threat level estimates via satellite imagery in contrast to witness estimates. In these types of situations, it is extremely beneficial that the consensus protocols are capable of accounting for this type of information and being able to “drive” or “force” the generated consensus towards them. Hence, we emphasize on several properties to be mandated by consensus protocols in order to generate a meaningful, or in other words, a rational consensus.

Thus, in order to be able to generate a rational consensus, *the consensus being attained must:*

\mathcal{P}_1 : be a weighted average of opinions among the agents [92];

\mathcal{P}_2 : reach the GT when it is known (i.e., all evidence must converge to GT); and

\mathcal{P}_3 : be “consistent” with a fully reliable estimate of the GT when it is available (i.e., the consensus must be “contained” within the estimate of the GT).

6.3.1 Consensus in a DST Fusion Environment

Consider a soft/hard data fusion environment, where consensus is sought among a set of m agents who maintain their current state as a DST BoE. We are now interested in developing a consensus protocol that is applicable to such scenarios. Let us proceed by formally defining the consensus in this context.

Definition 32 (DST Consensus). *Consider a set $\mathcal{N} = \{\mathcal{A}_1, \dots, \mathcal{A}_m\}$ of agents embedded at each discrete event-based index k in a directed graph $G[k] = (\mathcal{N}, E[k])$, where every edge $e_{ij} \in E[k]$ represents a unidirectional information exchange link from agent \mathcal{A}_i to agent \mathcal{A}_j . State of agent \mathcal{A}_i at discrete event-based index k is given by a BoE $\mathcal{E}_{\Theta_i}[k] \equiv \{\Theta, \mathfrak{F}_{\Theta_i}[k], m_{\Theta_i}(\cdot)[k]\}$, for $i = 1, \dots, m$.*

Each agent $\mathcal{A}_i \in \mathcal{N}$ starts with an initial state $\mathcal{E}_{\Theta_i}[0]$ and repeatedly update their states (via a valid DST updating strategy) by exchanging information according to the communication network specified by $G[k]$. Then, we say a consensus is reached among agents in \mathcal{N} , if $\|\mathcal{E}_{\Theta_i}[k] - \mathcal{E}_{\Theta_j}[k]\| \rightarrow 0$ as $k \rightarrow \infty$ for all $\mathcal{A}_i, \mathcal{A}_j \in \mathcal{N}$, for some DST norm $\|\cdot\|$. □

The desired properties \mathcal{P}_1 - \mathcal{P}_3 of a rational consensus can be translated into the DST notions as follows (see Figure 6.6 for a graphical illustration).

Definition 33 (Rational Consensus). Let $\mathcal{E}_\Theta^t \equiv \{\Theta, \mathfrak{F}_\Theta^t, m_\Theta^t(\cdot)\}$ and $\hat{\mathcal{E}}_\Theta^t \equiv \{\Theta, \hat{\mathfrak{F}}_\Theta^t, \hat{m}_\Theta^t(\cdot)\}$ denote the GT and a reliable estimate of GT, respectively. Also, let the BoE $\mathcal{E}_\Theta^* \equiv \{\Theta, \mathfrak{F}_\Theta^*, m_\Theta^*(\cdot)\}$ denote a consensus reached by the agents in \mathcal{N} via some valid consensus protocol. Now, we say \mathcal{E}_Θ^* is rational if

- i). \mathcal{E}_Θ^* is generated via a weighted average of state BoEs of agents in \mathcal{N} ;
- ii). \mathcal{E}_Θ^* reaches \mathcal{E}_Θ^t when it is known; and
- iii). \mathcal{E}_Θ^* is a refinement of $\hat{\mathcal{E}}_\Theta^t$ when it is known, in the sense that for all propositions $B \in \mathfrak{F}_\Theta^*$, $B \subseteq C \in \mathfrak{F}_\Theta^t$.

□

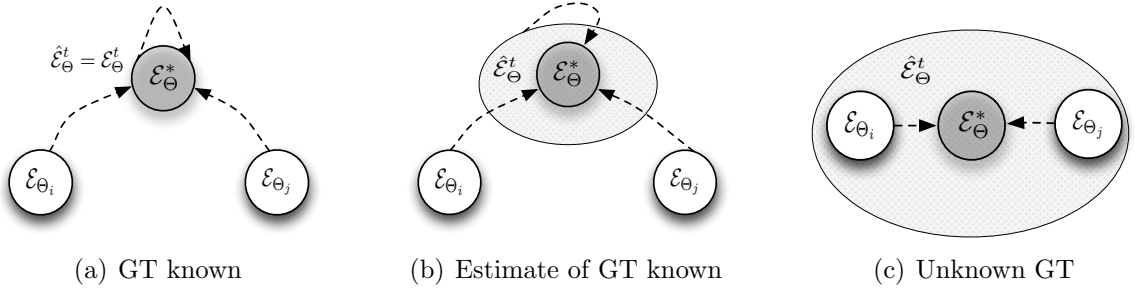


Figure 6.6: A Rational Consensus in DST Sense

What is required for the task at hand is an appropriate DST evidence updating strategy that generates a rational consensus, i.e., a consensus that is compliant with Definition 33. As we have discussed in Chapter 5, the CUE obviously makes sense in this situation and probably the most suitable candidate, given its flexibility in parameter selection and most importantly robustness against contradictory evidence—which is unavoidable in consensus scenarios¹⁷. However, we need to address some important questions before diving into the task of developing consensus protocols.

¹⁷Note that in most cases the BoEs cannot be discounted, simply because their reliability is unknown. In fact, the consensus can actually be used to estimate the credibility of a given BoE by comparing it against the consensus. see Chapter 7 for details.

- Q1.** Is it possible to establish a consensus protocols that generates a rational consensus in the sense of Definition 33 ?
- Q2.** What is an appropriate updating strategy ? Can we achieve a consensus by using the CUE ? Can the convergence be theoretically established ?
- Q3.** Can we establish consensus in partially-connected sensor networks (e.g., distributed networks) ?
- Q4.** How can we establish consensus protocols for networks with dynamic links (e.g., ad-hoc networks) and communication impairments (i.e., asynchronous consensus protocols) ?

We provide answers to all of the questions posted above with a single unified solution via setting up the consensus problem as an asynchronous iteration problem.

6.3.2 Asynchronous Consensus Protocols for DST BoEs

In this section, we establish a DST consensus protocol that achieve a rational consensus in the sense of Definition 33. We make use of the machinery from the previous section to generate convergence criteria that are applicable to various network topologies.

6.3.2.1 A CUE-based Paracontraction

We propose a particular CUE parameter selection that achieves several desirable properties for the given context. As it turns out, a consensus generated via this parameter selection satisfies all the properties \mathcal{P}_1 - \mathcal{P}_3 , and hence compliant with Definition 33 (see Claim 18).

Definition 34. Let $\mathfrak{E}_\Theta \equiv \{\mathcal{E}_\Theta \mid \mathcal{E}_\Theta = \{\Theta, \mathfrak{F}_\Theta, m_\Theta(\cdot)\}\}$ denote the set of all possible BoEs defined on Θ . Now, consider the set of n BoEs¹⁸, $\mathcal{E}_{\Theta_i} \in \mathfrak{E}_\Theta$, $i = 1, \dots, n$. Then, the operator $F_\triangleleft^i : \mathfrak{E}_\Theta^n \mapsto \mathfrak{E}_\Theta$ that updates \mathcal{E}_{Θ_i} with all \mathcal{E}_{Θ_j} s.t. $j \neq i$ is defined as

$$F_\triangleleft^i(\mathcal{E}_{\Theta_1}, \dots, \mathcal{E}_{\Theta_n}) \equiv \mathcal{E}_{\Theta_i} \triangleleft (\mathcal{E}_{\Theta_1} \bowtie \dots \bowtie \mathcal{E}_{\Theta_{i-1}} \bowtie \mathcal{E}_{\Theta_{i+1}} \bowtie \dots \bowtie \mathcal{E}_{\Theta_n}), \quad (6.13)$$

where the CUE parameters are given by

$$\alpha_i = C_i, \quad (6.14)$$

$$\beta_{ij}(A) = \begin{cases} C_j m_i(A), & \text{for } \mathcal{E}_i \equiv \hat{\mathcal{E}}^t; \\ C_j m_j(A), & \text{otherwise,} \end{cases} \quad (6.15)$$

such that $\alpha_i + \sum_{j \in \{1, \dots, n\} \setminus i} \sum_{A_{ij} \in \mathfrak{F}_{\Theta_j}} \beta_{ij}(A_{ij}) = 1$, where $C_i \in \mathbb{R}^+$, $i = 1, \dots, n$. \square

Remarks:

- Here, the parameter selections are based on rCUE and cCUE strategies.
- As one would have expected, notice that the parameter selection for an agent, who is not providing an estimate of the GT, is in fact rCUE. This allows for the agent to be “receptive” to the other agents. On the other hand, an agent providing a very reliable estimate has to be more “cautious,” and hence the cCUE strategy. This is exactly what guarantees a refinement only on the initial BoE of such an agent.
- The parameters C_i , $i = 1, \dots, m$ allows to incorporate relative “importance” factors of agents, when such information is available.

¹⁸The set of BoEs is taken to contain an estimate of GT when it exists, i.e., if $\exists \hat{\mathcal{E}}^t$, then $\mathcal{E}_{\Theta_i} \equiv \hat{\mathcal{E}}^t$ for some \mathcal{E}_{Θ_i} . Note that only one such estimate is allowed in the consensus process.

- Note, only one $\hat{\mathcal{E}}_{\Theta}^t$ is allowed to be incorporated into the consensus process; and obviously there can only be one fully reliable estimate of GT.

Claim 16. *The operator $F_{\triangleleft}^i : \mathfrak{E}_{\Theta}^n \mapsto \mathfrak{E}_{\Theta}$ as given in Definition 34 has infinitely many fixed points. Furthermore, let $\mathcal{E}_{\Theta} \in \text{fix}(F_{\triangleleft}^i)$ be arbitrary. Then, for all $B \in \mathfrak{F}_{\Theta}$, $\nexists C \in \mathfrak{F}_{\Theta}$ s.t. $B \subset C$ or $B \supset C$.* \square

Proof. see Section A.3.1 in Appendix A. \blacksquare

In other words, the Claim 16 establishes the fact that the operator F_{\triangleleft} can be used to generate fixed-points in \mathfrak{E}_{Θ} . Further, it also states that the fixed-points (BoEs) will have a certain structure; for any focal element B of a given fixed-point (BoE), there won't be any other focal element that containing or contained in B (see Figure 6.6 for a set theoretic interpretation). One can interpret this as fixed-points containing propositions that do not allow further “refinements.”

With this, we now make a claim that sets the foundation to develop consensus protocols based on CUE.

Claim 17. *The operator $F_{\triangleleft}^i : \mathfrak{E}_{\Theta}^n \mapsto \mathfrak{E}_{\Theta}$ as given in Definition 34 is a paracontractive on \mathfrak{E}_{Θ} w.r.t. any p -norm, $\|\cdot\| : \mathfrak{E}_{\Theta} \mapsto \mathbb{R}$ given by $\|\mathcal{E}_{\Theta}\| = \left(\sum_{B \subseteq \Theta} |m_{\Theta}(B)|^p \right)^{\frac{1}{p}}$.* \square

Proof. See Section A.3.2 in Appendix A. \blacksquare

After identifying the above CUE-based updating scheme as a paracontraction on the domain of DST BoEs, we are now in a position to define a consensus protocol for DST fusion environment.

Definition 35 (Asynchronous DST Consensus Protocol). *Let $t_0 < t_1 < \dots < t_k < \dots$ denote the event-based discrete time indices that the state of multi-agent system undergoes changes. Let the BoE $\mathcal{E}_{\Theta_i}[k]$ denote the state of agent \mathcal{A}_i at time t_k for $k = 0, 1, \dots$. Let $\mathcal{Q} = \{Q_{i,j} \mid j = 1, \dots, n_i, \quad i = 1, \dots, m\}$ denote the set of all*

interaction topologies observed during the consensus process. Let $J[k] \in \mathcal{Q}$ identify the agent interactions at time t_k , for $k = 0, 1, \dots$. Then, an asynchronous consensus protocol is given by

$$\mathcal{E}_{\Theta_i}[k+1] = \begin{cases} F_{\triangleleft}^{i_j}(\mathcal{E}_{\Theta_1}[s^1(k)], \dots, \mathcal{E}_{\Theta_{m_{I[k]}}}[s^{m_{I[k]}}(k)]), & \text{if } i = I[k]; \\ \mathcal{E}_{\Theta_i}[k], & \text{if } i \neq I[k], \end{cases} \quad (6.16)$$

where, $F_{\triangleleft}^{i_j}$ specifies the updating corresponding to $Q_{i,j}$ and $s^\ell[k] \in \mathbb{N}_0$ s.t. $s^\ell[k] \leq k$ for $\ell = 1, \dots, m$, $k = 0, 1, \dots$. \square

Remarks:

- This protocol allows for CUE-based evidence updating for agent BoEs in an asynchronous fashion.
- Agent interaction can be via fully or partially connected networks that are either static or dynamic.

Following is a very important claim that justifies the use of the above protocol in soft/hard fusion environment.

Claim 18. *A consensus BoE generated using the protocol in Definition 35 is a rational consensus in accordance with Definition 33.* \square

6.3.3 Existence of Consensus: Proof of Convergence

We are now in a position to derive the convergence of the consensus protocol in Definition 35. First derive the equivalent asynchronous iteration $(\mathcal{F}_{\triangleleft}, \mathcal{Y}_{\mathcal{O}}, \mathcal{I}, \mathcal{S})$ following the steps **Step**⁽¹⁾ – **Step**⁽⁴⁾ as described in Section 6.2.5.1, where $\mathcal{F}_{\triangleleft} = \{F_{\triangleleft}^i \mid i \in \mathbb{I}\}$, $\mathcal{Y}_{\mathcal{O}} \equiv \{y[-k] \in \mathfrak{E}_{\Theta} \mid y[-k] = \mathcal{E}_{\Theta_k}[0]\mathbf{1}, \text{ for } k = 1, \dots, m\}$ and $\mathcal{I} = \{I[k] \in \mathbb{I} \mid k = 0, 1, \dots\}$, where $\mathbb{I} = \{1, \dots, m\}$. We now claim that

Claim 19. *The pool $\mathcal{F}_{\triangleleft}$ is paracontractive on \mathfrak{E}_{Θ} and contains infinitely many common fixed-points in \mathfrak{E}_{Θ} .* \square

Proof. see Section A.3.3 in Appendix A. \blacksquare

With the above in place, all we need to show is that the iteration $(\mathcal{F}_{\triangleleft}, \mathcal{Y}_{\mathcal{O}}, \mathcal{I}, \mathcal{S})$ is confluent in order to prove its convergence, thus establishing the convergence of the DST consensus protocol in Definition 35. Let us now study the existence of consensus (i.e., convergence) of the above protocol under several interaction topologies.

6.3.3.1 Synchronous Fully-connected Network

A fully-connected synchronous network represents perhaps the simplest agent setup, where each agent is connected to all the other agents and information is exchanged without any iteration delay (i.e., $k - s^j[k] = 0$). In this case, consensus protocol in equation (6.16) reduces to

$$\mathcal{E}_{\Theta_i}[k+1] = F_{\triangleleft}^i(\mathcal{E}_{\Theta_1}[k], \dots, \mathcal{E}_{\Theta_m}[k]), \quad i = 1, \dots, m, \quad k \geq 1. \quad (6.17)$$

Claim 20. *Consensus protocol as given in (6.17) converges in a fully-connected synchronous network.* \square

Proof. Let us first prove that the equivalent iteration $(\mathcal{F}_{\triangleleft}, \mathcal{Y}_{\mathcal{O}}, \mathcal{I}, \mathcal{S})$ of (6.17) is confluent. Clearly, the sequence of iteration graphs $G[k]$, $k \geq 1$ is repeatedly jointly rooted. (i) all the agents \mathcal{A}_i , $i = 1, \dots, m$ use its latest state for the update; (ii) $\mathbb{I} = \{1, \dots, m\}$ and \mathcal{I} is regulated¹⁹; and (iii) $k - s^\ell[k] = 0 \leq s$, $\forall k \in \mathbb{N}_0$, $\ell = 1, \dots, m$ for all $s \in \mathbb{N}_0$. Hence, from Lemma 15, it follows that $(\mathcal{F}_{\triangleleft}, \mathcal{Y}_{\mathcal{O}}, \mathcal{I}, \mathcal{S})$ is confluent.

Now, according to Claim 19 and Theorem 14, the iteration $(\mathcal{F}_{\triangleleft}, \mathcal{Y}_{\mathcal{O}}, \mathcal{I}, \mathcal{S})$ converges. Hence, the consensus protocol in (6.17) converges. \blacksquare

¹⁹ \mathcal{I} regulated guarantees the fact that all the elements in \mathbb{I} are used within a finite time span. In this case $I[k] = \mathbb{I}$ since all the agents update their states at all k , which can also be renumbered s.t. $I[k] \cup I[k+1] \cup \dots \cup I[k+m-1] = \mathbb{I}$.

6.3.3.2 Synchronous, Static, Partially-connected Network

A static partially-connected synchronous network represents a setup, where agents communicate without iteration delays (i.e., $k - s^j[k] = 0$) with a network topology that is not fully connected and that does not change over time. Therefore, some of the agents cannot communicate to other agents (see Figure xxx). In this case, consensus protocol in equation (6.16) reduces to

$$\mathcal{E}_{\Theta_i}[k+1] = F_{\triangleleft}^i(\mathcal{E}_{\Theta_1}[k], \dots, \mathcal{E}_{\Theta_{m_i}}[k]), \quad i = 1, \dots, m, \quad k \geq 1, \quad (6.18)$$

where $m_i \in \{1, \dots, m\}$.

Claim 21. *Consensus protocol as given in (6.18) converges as long as the graph union of interaction networks of all agents is connected.* □

Proof. Consider the equivalent iteration $(\mathcal{F}_{\triangleleft}, \mathcal{Y}_{\mathcal{O}}, \mathcal{I}, \mathcal{S})$ of (6.18). Since, graph union of interaction topology of agents is connected and each agent uses his own latest state for update, the interaction graphs $G[k]$ are repeatedly jointly rooted. Similar to the case of synchronous fully-connected network, clearly assumptions (i) and (iii) of Lemma 15 are satisfied. Now, $\mathbb{I} = \{1, \dots, m\}$, since the topology is static and the network is connected. Hence, \mathcal{I} is regulated. Hence, $(\mathcal{F}_{\triangleleft}, \mathcal{Y}_{\mathcal{O}}, \mathcal{I}, \mathcal{S})$ is confluent according to Lemma 15. Thus, the confluent iteration $(\mathcal{F}_{\triangleleft}, \mathcal{Y}_{\mathcal{O}}, \mathcal{I}, \mathcal{S})$ converges according to the Claim 19 and Theorem 14. Hence, the consensus protocol in (6.18) converges. ■

6.3.3.3 Synchronous, Dynamic, Partially-connected Network

A synchronous dynamic partially-connected synchronous network represents a setup, where agents communicate without iteration delays (i.e., $k - s^j[k] = 0$), and with a network topology that is not fully connected and changes over time. Therefore, the agents do not interact with all the other agents; and also their interaction topology

changes over time (see Figure 6.3). In this case, consensus protocol in equation (6.16) reduces to

$$\mathcal{E}_{\Theta_i}[k+1] = F_{\triangleleft}^{I[k]}(\mathcal{E}_{\Theta_1}[k], \dots, \mathcal{E}_{\Theta_{m_{I[k]}}}[k]), \quad i = 1, \dots, m, \quad k \geq 1, \quad (6.19)$$

where $m_i \in \{1, \dots, m\}$. Given that $\mathcal{F}_{\triangleleft}$ is paracontracting on \mathfrak{E}_{Θ} , we only need to satisfy confluence conditions on equivalent iteration $(\mathcal{F}_{\triangleleft}, \mathcal{Y}_{\mathcal{O}}, \mathcal{I}, \mathcal{S})$ in order to prove convergence. In a synchronous network, assumption (iii) of Lemma 15, viz., $k - s^{\ell}(k) \leq s, \forall k \in \mathbb{N}_0, \ell = 1, \dots, m$, for any $s \in \mathbb{N}_0$ is clearly satisfied. Clearly, assumption (i) of Lemma 15, is also satisfied. Thus, if the interaction sequence \mathcal{I} is regulated (i.e., one should be able to find a finite time span at any given time t_k , on which all the agents participate in the consensus process), while the interaction graphs $G[k], k \geq 1$ of the network are repeatedly jointly rooted, the consensus protocol (6.19) converges.

6.3.3.4 Asynchronous Fully-connected Network

A fully-connected asynchronous network represents an agent communication setup, where each agent is connected to all the other agents, but the information exchange is not synchronized (or delayed) (i.e., $k - s^j[k] < 0$). In this case, consensus protocol in equation (6.16) reduces to

$$\mathcal{E}_{\Theta_i}[k+1] = F_{\triangleleft}^i(\mathcal{E}_{\Theta_1}[s^1(k)], \dots, \mathcal{E}_{\Theta_m}[s^m(k)]), \quad i = 1, \dots, m, \quad k \geq 1. \quad (6.20)$$

Claim 22. *The consensus protocol as given in Equation 6.20 converges if the iteration delays are finite.* □

Proof. Consider the equivalent iteration $(\mathcal{F}_{\triangleleft}, \mathcal{Y}_{\mathcal{O}}, \mathcal{I}, \mathcal{S})$ of (6.18). Since the network topology is fully connected, the interaction graphs $G[k], k \geq 1$ are repeatedly jointly rooted. Clearly, $(\mathcal{F}_{\triangleleft}, \mathcal{Y}_{\mathcal{O}}, \mathcal{I}, \mathcal{S})$ satisfies the assumptions (i) and (ii) of Lemma 15. Thus, if the iteration delays are finite, viz., $k - s^{\ell}(k) \leq \infty \forall k \in \mathbb{N}_0, \ell = 1, \dots, m$,

then according to Lemma 15, the iteration $(\mathcal{F}_{\triangleleft}, \mathcal{Y}_{\mathcal{O}}, \mathcal{I}, \mathcal{S})$ is confluent. Then, the convergence is guaranteed via Claim 19 and Theorem 14. Hence, the consensus protocol in (6.20) converges. ■

This discussion can also be extended to study asynchronous, static and partially-connected networks. In this case, the similar to the case of static, partially-connected networks, one needs to impose conditions on how agents interact in order to make sure that there's adequate coupling among agents to reach a consensus. This can be setup by making sure that graph union of interaction topology of each agent is connected graph. In order to guarantee the confluence in iterations, one also needs to make sure that the iteration delays is finite at any given time.

6.3.3.5 An Arbitrary Network Topology

As we have shown (with theoretical proofs) with several network topologies, one needs to make sure that the coupling among agents is adequate to reach a consensus. Given that the CUE-based operators is paracontractive on on DST BoEs, the consensus protocol we have presented is capable to generating a rational consensus in any network setup as long as these coupling conditions are satisfied. In the analysis of an arbitrary network topology, one only need to make sure that it satisfies Lemma 15; and the convergence is automatically guaranteed via Theorem 14.

Note:

We illustrate the proposed consensus protocol by applying it to a credibility estimation, where the GT is estimated by consensus. See Chapter 7 for details.

6.4 Chapter Summary

Consensus analysis is a fundamentally important problem soft/hard fusion networks. We have presented a consensus protocol based on DST notions of evidence updating. Therefore the proposed method can be applied in soft/hard fusion envi-

ronments containing various types of data uncertainties and it is theoretically proven to generate a rational consensus. Convergence criteria is derived in such a manner, so that the same principles can be applied to analyze a wider class of network topologies. We have proven the convergence of the proposed consensus protocol under several network topologies; and provided criteria for convergence on arbitrary networks. The theoretical results are highlighted and confirmed by the observations made in the experiments. In the next Chapter, we make use of the notions of consensus in order to assess the credibility of sources in the absence of GT.

Data Validation for Robust Fusion

DUE to its subjective nature which can otherwise compromise the integrity of the fusion process, it is critical that soft data be validated prior to its incorporation into the fusion engine. The strategy of discounting evidence based on source reliability may not be applicable when dealing with soft sources because their reliability is often unknown beforehand. Thus, it is necessary that some mechanism is employed to validate the data when their credibility is unknown beforehand. Here, we propose a methodology based on the notion of consensus to estimate the credibility of (soft) evidence in the absence of GT. This estimated credibility can then be used for source reliability estimation, discounting or appropriately “weighting” evidence for fusion. We illustrate several interesting and intuitively appealing properties of the proposed method via a numerical example.

This chapter is organized as follows: Section 7.1 provides an overview of the data validation problem in soft/hard fusion domains; Section 7.2 describes the credibility estimation problem and presents a new strategy based on consensus; Section 7.3 highlights various properties of the proposed method and illustrates the consensus-based credibility estimation strategy; and finally, Section 7.4 contains the chapter summary.

7.1 Overview

Credibility refers to “... *the quality of being trusted and believed in [26].*” Thus, one can interpret credibility as an instantaneous measure of trustworthiness of evidence. Therefore, credibility of evidence, when provided, can be used to validate evidence prior to fusion. However, this information is not available in many applications; and often other relevant contextual/meta data one may use to estimate credibility are also not available in most circumstances. Reliability, which refers to a notion of “... *[being] consistently good in quality/performance or able to be trusted [26]*”, provides an overall performance figure of a source. This information can be used to “discount” for evidence whose credibility information is not known in order to guarantee a robust fusion.

7.2 Credibility Estimation

Discounting based on source reliability is not feasible in many, especially soft/hard, applications. A good example would be a situation assessment scenario, where a group of civilians are providing “eye-witness” statements to identify a suspicious vehicle: in such scenarios, this type of evidence often provide perhaps the most crucial information, but their credibility or the reliability of sources are often not known beforehand.

In typical soft/hard fusion scenarios, such as the above, soft evidence is usually gathered from many sources. Even though, the actual reliability of these sources are not fully known, they are usually not totally unreliable. Hence, it is not unreasonable to assume that the truth is reflected in the “aggregate of their opinions,” if an adequate number of sources are considered.

7.2.1 Conflict-based Credibility Estimation

A method to self-validate the evidence is provided in [93] by estimating the conflict among evidence. Here, the conflict of a body of evidence is “inversely” related to its credibility²⁰.

Definition 36. *Given the BoEs \mathcal{E}_{Θ_i} , $i = 1, \dots, n$, the credibility of \mathcal{E}_{Θ_i} is given by*

$$Cr_{cf}(\mathcal{E}_{\Theta_i}) = (1 - \text{conf}(\mathcal{E}_{\Theta_i}, \mathcal{E}_{j \neq i})^\lambda)^{1/\lambda}, \quad (7.1)$$

where $\lambda \in \mathbb{R}^+$, $\mathcal{E}_{j \neq i} = \{\mathcal{E}_{\Theta_j}, j = 1, \dots, n \mid j \neq i\}$ and $\text{conf}(\mathcal{E}_{\Theta_i}, \mathcal{E}_{j \neq i})$ is the conflict between \mathcal{E}_{Θ_i} and $\mathcal{E}_{j \neq i}$. Two variants Cr_{cf1} and Cr_{cf2} are

$$\text{conf}(\mathcal{E}_{\Theta_i}, \mathcal{E}_{j \neq i}) = \begin{cases} \frac{1}{n-1} \sum_{\substack{j=1; \\ j \neq i}}^n \text{dist}(\mathcal{E}_{\Theta_i}, \mathcal{E}_{\Theta_j}), & \text{for } Cr_{cf1}; \\ \text{dist}(\mathcal{E}_{\Theta_i}, \mathcal{E}_{\oplus j \neq i}), & \text{for } Cr_{cf2}, \end{cases}$$

where $\mathcal{E}_{\oplus j \neq i} = \mathcal{E}_{\Theta_1} \oplus \dots \oplus \mathcal{E}_{\Theta_{i-1}} \oplus \mathcal{E}_{\Theta_{i+1}} \oplus \dots \oplus \mathcal{E}_{\Theta_n}$, for $i = 1, \dots, n$ and $\text{dist}(\cdot, \cdot)$ is a distance measure for DST BoEs as given in Definition 10. \square

7.2.2 Consensus-based Credibility Estimation

With credibility viewed as a measure of the instantaneous trustworthiness of evidence, it makes sense to assess the credibility of a BoE by comparing it to the GT via a distance measure (such as what appears in Definition 10 in Chapter 3):

Definition 37. *Let \mathcal{E}^t denote the GT. Then, the credibility of the BoE \mathcal{E} is given by*

$$Cr_{con}(\mathcal{E}) = (1 - \text{dist}(\mathcal{E}, \mathcal{E}^t)^\lambda)^{1/\lambda}, \text{ where } \lambda \in \mathbb{R}^+.$$

²⁰The authors in [93] refer to this as a measure of relative reliability. However, to be consistent with our interpretations of the terms, we take their definition as a measure of credibility.

As sensible as it appears, the difficulty with this strategy lies in the fact that the GT is usually absent. Is there a way to estimate the GT in such a situation? The notion of consensus has been used in many disciplines (e.g., social sciences, marketing/finance, engineering) and in a myriad of applications as a method to arrive at a “general agreement” among opinions or sources. Here, we make use of the DST consensus protocol that was developed in Chapter 6 to generate an estimate of GT from the available evidence. This strategy provides an ideal method to estimate the GT due to some of its properties; for instance the consensus BoE $\mathcal{E}_\Theta^* \equiv \{\Theta, \mathfrak{F}_\Theta^*, m_\Theta^*(.)\}$ will always be a refinement of a reliable estimate of GT fed into the system $\hat{\mathcal{E}}_\Theta^t \equiv \{\Theta, \hat{\mathfrak{F}}_\Theta^t, \hat{m}_\Theta^t(.)\}$.

7.3 An Illustrative Example

Consider a multi-agent system that consists of 5 agents $\mathcal{A}_i, i = 1, \dots, 5$, where each agents evidence is given by $\mathcal{E}_{\Theta_i} = \{\Theta \equiv \{abcde\}, \mathfrak{F}_{\Theta_i}, m_{\Theta_i}(.)\}$, for $i = 1, \dots, 5$. Suppose their reliabilities are unknown. We also have a 100% reliable estimate of GT given by $\hat{\mathcal{E}}_\Theta = \{\Theta, \hat{\mathfrak{F}}_\Theta^t, \hat{m}_\Theta^t(.)\}$

Setup: Suppose the BPAs are as follows:

$$\begin{aligned} m_{\Theta_1}(ac)=0.9; \quad m_{\Theta_2}(b) &=0.9; \quad m_{\Theta_3}(ac)=0.9; \quad m_{\Theta_4}(ac)=0.9; \quad m_{\Theta_5}(e) &=0.9; \\ m_{\Theta_1}(b) &=0.1; \quad m_{\Theta_2}(abc)=0.1; \quad m_{\Theta_3}(e) &=0.1; \quad m_{\Theta_4}(d) =0.1; \quad m_{\Theta_5}(abc)=0.1. \end{aligned}$$

We consider four cases (in decreasing order of “preciseness” of the GT estimate):

Case 1	Case 2	Case 3	Case 4
$\hat{m}_\Theta^t(a)=1.0;$	$\hat{m}_\Theta^t(ab)=1.0;$	$\hat{m}_\Theta^t(abc)=1.0;$	$\hat{m}_\Theta^t(\Theta)=1.0.$

The cases simulates the following scenarios:

Case 1: GT is known

Cases 2-3: only an estimate of the GT is known; and

Case 4: the GT is completely unknown.

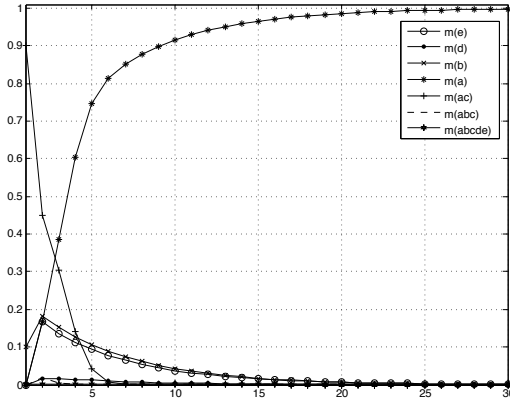
For each case, all the six BoEs \mathcal{E}_{Θ_i} , $i \in 1, \dots, 5$, reach the following consensus BoE:

Case 1: $m_{\Theta}^*(a) = 1.00$

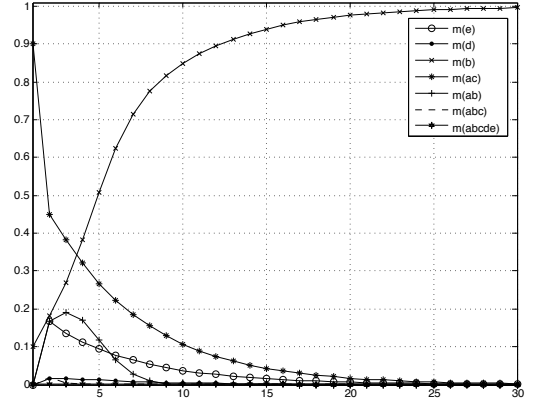
Case 2: $m_{\Theta}^*(b) = 1.00$

Case 3: $m_{\Theta}^*(b) = 0.29$ $m_{\Theta}^*(ac) = 0.71$

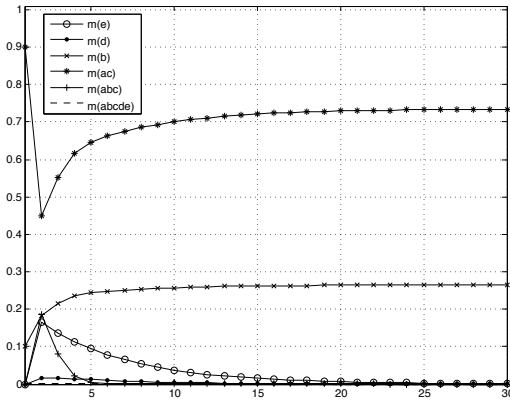
Case 4: $m_{\Theta}^*(b) = 0.30$ $m_{\Theta}^*(ac) = 0.66$ $m_{\Theta}^*(d) = 0.02$ $m_{\Theta}^*(e) = 0.02$



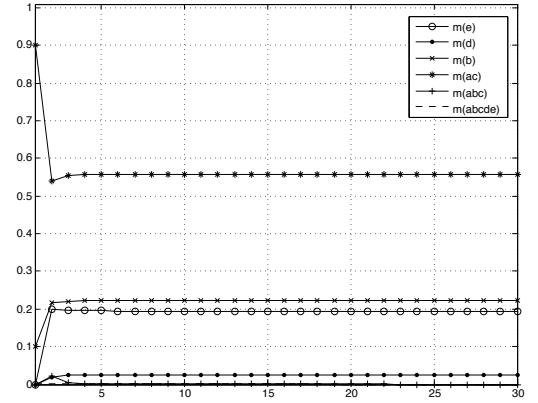
(a) Case 1: $\hat{m}_{\Theta}^t(a) = 1.0$



(b) Case 2: $\hat{m}_{\Theta}^t(ab) = 1.0$



(c) Case 3: $\hat{m}_{\Theta}^t(abc) = 1.0$



(d) Case 4: $\hat{m}_{\Theta}^t(\Theta) = 1.0$

Figure 7.1: Convergence of \mathcal{E}_{Θ_1} to \mathcal{E}_{Θ}^* as indicated by the evolution of the BPA with k . All the focal elements that are not contained in the core of the estimated GT $\hat{\mathcal{E}}_{\Theta}^t$ vanish as \mathcal{E}_{Θ_1} reaches \mathcal{E}_{Θ}^* .

Figure 7.1 shows the convergence of BoE \mathcal{E}_{Θ_1} to \mathcal{E}_{Θ}^* for each case. Note how the consensus BoE is ‘consistent’ or ‘agrees’ with \mathcal{E}_{Θ}^t . Behavior of other BoEs are similar and converge to \mathcal{E}_{Θ}^* in each case.

Method	Credibility																			
	Case 1. $\hat{m}_{\Theta}^t(a) = 1.0$					Case 2. $\hat{m}_{\Theta}^t(ab) = 1.0$					Case 3. $\hat{m}_{\Theta}^t(abc) = 1.0$					Case 4. $\hat{m}_{\Theta}^t(\Theta) = 1.0$				
	\mathcal{E}_{Θ_1}	\mathcal{E}_{Θ_2}	\mathcal{E}_{Θ_3}	\mathcal{E}_{Θ_4}	\mathcal{E}_{Θ_5}	\mathcal{E}_{Θ_1}	\mathcal{E}_{Θ_2}	\mathcal{E}_{Θ_3}	\mathcal{E}_{Θ_4}	\mathcal{E}_{Θ_5}	\mathcal{E}_{Θ_1}	\mathcal{E}_{Θ_2}	\mathcal{E}_{Θ_3}	\mathcal{E}_{Θ_4}	\mathcal{E}_{Θ_5}	\mathcal{E}_{Θ_1}	\mathcal{E}_{Θ_2}	\mathcal{E}_{Θ_3}	\mathcal{E}_{Θ_4}	\mathcal{E}_{Θ_5}
Cr_{cf1}	0.32	0.05	0.32	0.32	0.06	0.47	0.17	0.46	0.45	0.12	0.51	0.15	0.51	0.50	0.13	0.52	0.12	0.52	0.51	0.14
	5	1	5	5	2	5	2	4	3	1	5	2	5	3	1	5	1	5	3	2
Cr_{cf2}	0.32	0.05	0.32	0.32	0.06	0.32	0.05	0.32	0.32	0.06	0.90	0.07	0.90	0.90	0.08	0.90	0.07	0.90	0.90	0.08
	5	1	5	5	2	5	1	5	5	2	5	1	5	5	2	5	1	5	5	2
Cr_{con}	0.32	0.05	0.32	0.32	0.06	0.10	0.92	0.05	0.05	0.06	0.83	0.33	0.77	0.77	0.19	0.71	0.37	0.70	0.67	0.38
	5	1	5	5	2	4	5	2	2	3	5	2	4	4	1	5	1	4	3	2

Table 7.1: Estimated credibility measures of the BoEs.

Credibility Estimation: We now use the consensus BoE \mathcal{E}_{Θ}^* in place of \mathcal{E}_{Θ}^t in Definition 37 to get the credibility estimates Cr_{con} for each BoE. See Table 7.1 which also shows the two measures Cr_{cf1} and Cr_{cf2} in Definition 36. Ranked credibility values (lowest is ‘1’) are also indicated underneath each credibility value in Table 7.1.

In Case 1, not surprisingly, all measures produce identical results. In Case 2, the assignment of a low credibility to \mathcal{E}_{Θ_2} (supporting proposition b) by both Cr_{cf1} and Cr_{cf2} is surprising when GT is either a or b . Cr_{con} assigns a significantly higher credibility to \mathcal{E}_{Θ_2} relative to other BoEs. The assignment of low credibility to $\mathcal{E}_{\Theta_1}, \mathcal{E}_{\Theta_3}, \mathcal{E}_{\Theta_4}$ (mainly supporting a or b) also needs further investigation. The comparison is more difficult with decreasing exactness of the GT estimate, but they all seem to agree. Cases 3-4 are illuminating: **(a)** consensus BoE allocates higher supports for ac , which is in concordance with what a cursory glance at the BoEs reveals; **(b)** a is absent in the consensus because no BoE supports the singleton a ; **(c)** d and e are absent in Case 3 consensus because they are absent in the GT estimate.

7.4 Chapter Summary

The proposed consensus-based credibility estimation strategy addresses an important research question: *how can we validate evidence when the GT is unknown?* It can be used for purposes of (a) estimating the source reliability, (b) weighting the sources for fusion, or (c) discounting the BoEs. We have used the DST consensus protocol developed in Chapter 6. This is the major difference with conflict-based methods which are heavily dependent on the combination rule utilized (e.g., when the Cr_{cf2} employs the DCR, evidence conflicts can generate null results. To avoid this, we set $m_i(\Theta) = 0.0001$ (and deducted 0.0001 from the largest mass). However, the conflict-based measures may be computationally more efficient than the consensus-based approach. This important issue warrants further investigation.

Part III

COMPUTATIONAL OPTIMIZATIONS

Efficient Computation of DST Conditionals

CONDITIONING in DST, especially with the FH conditionals, carries a significant, sometimes even prohibitive, computational overhead. To quote [56], “... *the operation of conditioning can cause an exponential explosion in the number of nonzero Möbius assignments used to represent the function.*” Moreover, the computational complexity exponentially grows with increasing cardinality of the FoD [27]. As a result, DST methods involving conditionals can quickly become intractable in hard/soft data fusion scenarios where the FoDs of sources are often of high cardinality. However, the CCT proposed in Chapter 4 provides an appealing path that can be pursued towards the development of efficient algorithms for computing FH conditionals. In this chapter, a graphical structure and a message-passing scheme for efficiently computing the FH conditionals are developed. Bounds on computational complexity, conditions that would ensure the proposed algorithm to yield computational advantages, and extensive simulations to illustrate these advantages, are also provided.

This chapter is organized as follows: Section 8.1 provides a chapter overview addressing the issues related to conditional computations; Section 8.2 provides a method to directly compute the conditional masses by exploiting the CCT; Section 8.3 presents an analysis on computational gains and also provide criteria for selecting an

optimal computational method; Section 8.4 contains several experimental simulations; and finally Section 8.5 provides the chapter summary.

8.1 Chapter Overview

As Kennes, et al., mentions [27], as the size of the FoD is increased, conditioning causes an “exponential explosion.” This is simply due to the fact that many of the more commonly utilized DST conditional notions, including the FH conditionals, are defined in terms of either belief or plausibility [28, 69, 94, 95]. Therefore, one has to compute and store belief/plausibility values for essentially all the elements of the powerset of the FoD, For instance, an FoD having cardinality 40 has $2^{40} \approx 10^{12}$ belief values; with floating number representation, this requires approximately $8TB$ of storage space.

However, most real-world evidence sources allocate non-zero mass to only a few propositions. Thus, it is often computationally efficient to store and operate on the mass assignments instead of the alternate DST notions of belief or plausibility [96]. Since the CCT can directly identify the conditional focal elements, the need to compute the conditional beliefs/plausibilities of all the subsets of the conditioning event and then relying on a Möbius transformation to get the conditional masses is eliminated. Once the conditional focal elements are thus identified, by representing the conditional core as a directed graph, we present a simple uni-directional message passing scheme to compute the conditional masses. The conditional beliefs of the conditional focal elements are also generated as an intermediate result. The proposed algorithm has the potential to provide significant computational savings in larger FoDs where the computation and storage of belief/plausibility functions are prohibitive. We also derive specific conditions that identify the scenarios where the algorithm provides computational savings.

8.2 Direct Computation of Conditional Masses

Here we illustrate how conditional masses can be directly computed, once they are identified using the CCT. This requires the conditional focal elements to be “properly ordered,” so that the computations can be carried out without having to access the belief and/or plausibility functions. Let us proceed by analyzing the definition of FH conditionals.

Expand the belief terms expression for $\text{Bl}_\Theta(B|A)$ in Claim 3 in terms of mass functions to get

$$\sum_{C \subseteq B} m_\Theta(C|A) = \frac{\sum_{C \subseteq B} m_\Theta(C)}{\text{Pl}_\Theta(A) - \sum_{\substack{\emptyset \neq C \subseteq B; \\ \emptyset \neq D \subseteq \bar{A}}} m_\Theta(C \cup D)} \quad (8.1)$$

$$\Rightarrow m_\Theta(B|A) = \frac{\sum_{C \subseteq B} m_\Theta(C)}{\text{Pl}_\Theta(A) - \sum_{\substack{\emptyset \neq C \subseteq B; \\ \emptyset \neq D \subseteq \bar{A}}} m_\Theta(C \cup D)} - \sum_{C \subset B} m_\Theta(C|A) \quad (8.2)$$

We make the following observations:

1. $\text{Pl}_\Theta(A)$ is common to all $B \subseteq \Theta$.
2. Computation of $m(B|A)$ requires access to,
 - (a) all $C \in \mathfrak{F}_\Theta$ s.t. $C \subseteq B$;
 - (b) all $(C \cup D) \in \mathfrak{F}_\Theta$ s.t. $C \subseteq B$ and $D \subseteq \bar{A}$; and
 - (c) all $C \in \mathfrak{F}_{\Theta|A}$ s.t. $C \subset B$.
3. If the $m_\Theta(\cdot|A)$ computation is “properly ordered,” the availability of $\text{Bl}_\Theta(B)$ and $\text{Pl}_\Theta(B)$ for all $B \subseteq \Theta$ is not required.

With these observations in place, we propose to represent the conditional focal elements in a polytree as described below. Then, the decomposition in (8.2) can be used to derive a simple message passing scheme to compute the conditional masses.

8.2.1 Polytree Representation of Conditional Core

Let $\mathcal{X} = \{X_1, \dots, X_n\}$ denote $n = |\mathfrak{F}_{\Theta|A}|$ number of nodes. For $i = 1, \dots, n$, the conditional focal element $B_{X_i} \in \mathfrak{F}_{\Theta|A}$ is represented by the node X_i . The sets $\mathbf{Ch}(X_i)$ and $\mathbf{Pa}(X_i)$ represent the *child nodes* and *parent nodes* of node $X_i \in \mathcal{X}$. If $X_j \in \mathbf{Ch}(X_i)$, then it is represented by a *directed path or edge* $X_i \rightarrow X_j$, $i \neq j$. Note that

$$\mathbf{Ch}(X_i) = \{X \in \mathcal{X} \mid B_X = \text{Lgst} \{B \in \mathfrak{F}_{\Theta|A} \mid B \subset B_{X_i}\}\}. \quad (8.3)$$

If there is a directed path from node X_i to X_j , then we say that the node X_j is a *descendent* of X_i . The set $\mathbf{Des}(X_i)$ represents the set of descendent nodes of X_i ; we use $\mathbf{Des}^*(X_i)$ to represent $\mathbf{Des}(X_i) \cup \{X_i\}$. A node X_i for which $\mathbf{Des}(X_i) = \emptyset$ is referred to as a *leaf node*. The polytree corresponding to the conditional core $\mathfrak{F}_{\Theta|A}$ in Example 1 appears in Figure 8.1.

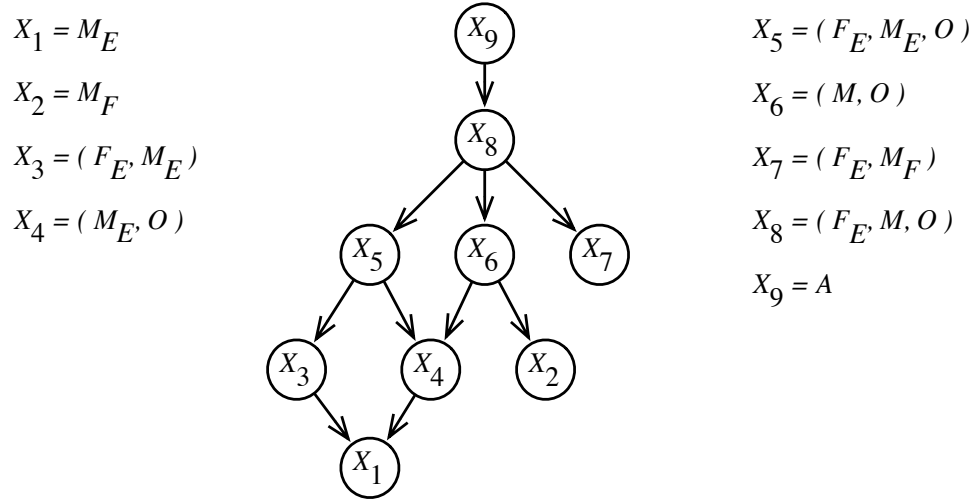


Figure 8.1: Polytree corresponding to $\mathfrak{F}_{\Theta|A}$ in Example 1.

Define the collections

$$\begin{aligned} \mathbf{S}_{X_i} &= \{B \in \mathfrak{F}_{\Theta} \mid (B \cap A) \in \text{out}(A), B \cap A \subseteq B_{X_i}, B \cap A \not\subseteq B_{X_j}, \forall X_j \in \mathbf{Des}(X_i)\}; \\ \mathbf{F}_{X_i} &= \{B \in \mathfrak{F}_{\Theta} \mid B \subseteq B_{X_i}, B \not\subseteq B_{X_j}, \forall X_j \in \mathbf{Des}(X_i)\}, \end{aligned} \quad (8.4)$$

to identify the focal elements that first appear in node X_i without appearing in its decendents. Let us also define

$$\mathcal{S}_{X_i}^{(F)} = \sum_{B \in \mathbf{F}_{X_i}} m_{\Theta}(B); \quad \mathcal{S}_{X_i}^{(S)} = \sum_{B \in \mathbf{S}_{X_i}} m_{\Theta}(B); \quad \mathcal{S}_{X_i}^{(|A)} = m_{\Theta}(B_{X_i}|A). \quad (8.5)$$

8.2.2 Conditional Masses via Message-Passing

The node X_i needs the messages $\mu_{X_j \rightarrow X_i}(X_i)$, $\forall X_j \in \mathbf{Ch}(X_i)$, to compute the conditional mass $m_{\Theta}(B_{X_i}|A)$. For node X_j to pass the message $\mu_{X_j \rightarrow X_i}(X_i)$ to a node $X_i \in \mathbf{Pa}(X_j)$, it requires $m_{\Theta}(B_{X_j}|A)$. Thus, the message passing is only *upstream*; each node $X_j \in \mathcal{X}$ passes the message $\mu_{X_j \rightarrow X_i}(X_i)$ to its parent node X_i , $\forall X_i \in \mathbf{Pa}(X_j)$. The computations are initiated at the *leaf nodes*. So, we arrive at the following message passing scheme:

1. Each child node X_j passes the following message matrix to its parent X_i :

$$\mu_{X_j \rightarrow X_i}(X_i) = \begin{bmatrix} \mathbf{Des}^*(X_j) & \mu_{X_j \rightarrow X_i}^{(F)} & \mu_{X_j \rightarrow X_i}^{(S)} & \mu_{X_j \rightarrow X_i}^{(|A)} \end{bmatrix}, \quad (8.6)$$

where

$$\mu_{X_j \rightarrow X_i}^{(\cdot)} = \begin{bmatrix} \mathcal{S}_{X_j}^{(\cdot)} & \mathcal{S}_{X_{k_1}}^{(\cdot)} & \dots & \mathcal{S}_{X_{k_n}}^{(\cdot)} \end{bmatrix}^T, \quad X_{k_\ell} \in \mathbf{Des}(X_j), \quad \ell = 1, \dots, n. \quad (8.7)$$

2. At each node $X_i \in \mathcal{X}$, the conditional mass $m_{\Theta}(B_{X_i}|A)$ is calculated as

$$m_{\Theta}(B_{X_i}|A) = \frac{\mathcal{S}_{X_i}^{(F)} + \sum_{X_j \in \mathbf{Des}(X_i)} \mathcal{S}_{X_j}^{(F)}}{\text{Pl}_{\Theta}(A) - \mathcal{S}_{X_i}^{(S)} - \sum_{X_j \in \mathbf{Des}(X_i)} \mathcal{S}_{X_j}^{(S)}} - \sum_{X_j \in \mathbf{Des}(X_i)} m_{\Theta}(B_{X_j}|A), \quad (8.8)$$

where the updated descendants vector of X_i is given by

$$\mathbf{Des}(X_i) \leftarrow \bigcup_{X_j \in \mathbf{Ch}(X_i)} \mathbf{Des}(X_j). \quad (8.9)$$

3. At any leaf node $X_i \in \mathcal{X}$, the computation of $m(B_{X_i}|A)$ simplifies to

$$m_{\Theta}(B_{X_i}|A) = \frac{\mathcal{S}_{X_i}^{(F)}}{\text{Pl}_{\Theta}(A) - \mathcal{S}_{X_i}^{(S)}}. \quad (8.10)$$

8.3 Computational Gains

The CCT precisely identifies the $|\mathfrak{F}_{\Theta|A}|$ number of conditional focal elements, thus eliminating the need for computing conditional masses of all these subsets of A . We can easily derive the following upper bound on the size of the conditional core in terms of the size of sets $\text{in}(A)$, $\text{out}(A)$ and $\text{OUT}(A)$.

Claim 23 (Upper Bound on $|\mathfrak{F}_{\Theta|A}|$). *Let $N_{\text{in}} = |\text{in}(A)|$, $N_{\text{out}} = |\text{out}(A)|$, and $N_{\text{OUT}} = |\text{OUT}(A)|$. Then, $|\mathfrak{F}_{\Theta|A}| \leq N_{\text{in}}(1 + N_{\text{OUT}}) \leq N_{\text{in}}2^{N_{\text{out}}}$. \square*

Proof. Let $\mathfrak{B} = \{B = X \cup Y \text{ s.t. } \emptyset \neq X \in \text{in}(A), \emptyset \neq Y \in \text{OUT}(A)\}$. Then,

$$\begin{aligned} |\mathfrak{F}_{\Theta|A}| &= |\text{in}(A) \cup \mathfrak{B}| = |\text{in}(A) \cup (\mathfrak{B} \setminus \text{in}(A))| \\ &= |\text{in}(A)| + |\mathfrak{B}| - |\mathfrak{B} \cap \text{in}(A)| \\ &\leq N_{\text{in}} + N_{\text{in}}N_{\text{OUT}} - |\mathfrak{B} \cap \text{in}(A)| \\ &\leq N_{\text{in}} + N_{\text{in}}N_{\text{OUT}} \leq N_{\text{in}} + N_{\text{in}}(2^{N_{\text{out}}} - 1). \end{aligned}$$

This establishes the claim. \blacksquare

Remarks:

- If $|\text{in}(A) \cap \mathfrak{B}| \ll N_{\text{in}}(1 + N_{\text{OUT}})$, then $|\mathfrak{F}_{\Theta|A}| \leq N_{\text{in}}(1 + N_{\text{OUT}})$ provides a tighter bound.
- Also, note that N_{OUT} could be significantly smaller than $2^{N_{\text{out}}} - 1$, since elements of $\text{OUT}(A)$ are not necessarily disjoint. Thus, $N_{\text{in}}2^{N_{\text{out}}}$ may lead to a very conservative bound if there is significant overlap in the elements of $\text{OUT}(A)$.

Example 16. *Let us consider Example 1 in Chapter 4 again. Here, use of the CCT leads to about a 70% reduction in the required number of conditional mass computations. The CCT can be used to obtain significant computational savings in*

situations where the sets involved are of higher cardinality, and N_{in} and N_{out} are significantly small compared to $2^{|A|}$ (e.g., when dealing with soft evidence). For instance, take $|\Theta| = 50$, $|A| = 20$, $N_{in} = 500$, $N_{out} = 8$, and $N_{OUT} = 200$. The number of conditional mass computations reduces from $2^{20} - 2$ to a maximum of $500 \times 201 \leq 500 \times 2^8 = 128000$ (corresponding to about 88% reduction). \square

While the CCT can be used to identify the conditional focal elements with no recourse to numerical computations, the identification step itself takes non-zero computational power. Indeed, with the CCT, the number of times the conditional mass has to be computed is typically much less; however, one also has to take into account the overhead associated with identifying conditional focal elements and extra burden associated with data representations and computations. Thus, the overall computational gains are highly dependent on the implementation, as well as the application.

To provide specific conditions that would guarantee computational gains with the use of the CCT, we will make use of the implementation specific constants in Table 8.1.

Notation	Description
$\mathcal{T}_{id.ioa}$	Average cpu time to identify elements of $\text{in}(A)$ and $\text{out}(A)$
$\mathcal{T}_{tr.bld}$	Average cpu time for processing $\mathfrak{F}_{\Theta A}$ to build the polytree
$\mathcal{T}_{ps.msg}$	Average cpu time per message
$\mathcal{T}_{bp.ohd}$	Average overhead (e.g., $\text{Bl}_{\Theta}(\cdot)$ etc.) cpu time without the CCT
$\mathcal{T}_{cp.std}$	Average cpu time to compute $m_{\Theta}(\cdot A)$ without the CCT
$\mathcal{T}_{cp.cct}$	Average cpu time to compute $m_{\Theta}(\cdot A)$ with the CCT
$\mathcal{T}_{tr.pro}$	$\mathcal{T}_{tr.bld} + \mathcal{T}_{ps.msg}$

Table 8.1: Some Implementation Specific Constants (on a per proposition basis)

Notice the following about the graph structure and the proposed message passing scheme for conditional mass computation:

1. The graph structure can be readily obtained while generating conditional focal elements by taking the unions of elements of $\text{in}(A)$ with $\text{OUT}(A)$.

2. One may start by setting the elements in $\text{in}(A)$ as leaf nodes, and adding their parent nodes by taking unions of leaf nodes with $\text{out}(A)$.
3. Since the message passing is up-stream only, mass computation and message passing can coincide. Thus, any node is accessed only once.

So, the total computational time can be estimated as follows:

CCT Method:

$$\mathcal{T}_{thm} = 2^{|\Theta|} \cdot \mathcal{T}_{id.ioa} + N_{in} 2^{N_{out}} \cdot \mathcal{T}_{tr.pro} + |\mathfrak{F}_{\Theta|A}| \cdot \mathcal{T}_{cp.cct};$$

Standard Method:

$$\mathcal{T}_{std} = 2^{|\Theta|} \cdot \mathcal{T}_{bp.ohd} + 2^{|A|} \cdot \mathcal{T}_{cp.std}. \quad (8.11)$$

Here, the *CCT method* refers to the method of using the CCT to directly identify the conditional core and then using the proposed graphical structure and the message passing scheme; the *standard method* refers to the direct method of Möbius transforming the conditional beliefs to obtain the conditional masses. Therefore, the CCT method is guaranteed to reduce the computational burden if $\mathcal{T}_{thm} < \mathcal{T}_{std}$. Once a BoE and a conditioning event are given, the following procedure can be used to systematically check whether the application of the CCT method is useful.

Lemma 24 (Criterion 1). *Evaluate N_{in} and N_{out} . This only takes $2^{|\Theta|}\mathcal{T}_{id.ioa}$ of computational time. Then check for*

$$N_{in} 2^{N_{out}} < \frac{2^{|\Theta|} \cdot (\mathcal{T}_{bp.ohd} - \mathcal{T}_{id.ioa}) + 2^{|A|} \cdot \mathcal{T}_{cp.std}}{\mathcal{T}_{tr.pro} + \mathcal{T}_{cp.cct}}.$$

If this is true, then the CCT is guaranteed to provide computational gains; if it fails, check Criterion 2. ■

Lemma 25 (Criterion 2). *Evaluate N_{OUT} . Then check for*

$$N_{in}2^{N_{out}} \cdot \mathcal{T}_{tr.pro} + N_{in}(1 + N_{OUT}) \cdot \mathcal{T}_{cp.cct} < 2^{|\Theta|} \cdot (\mathcal{T}_{bp.ohd} - \mathcal{T}_{id.ioa}) + 2^{|A|} \cdot \mathcal{T}_{cp.std}.$$

If this is true, then the CCT is guaranteed to provide computational gains. ■

If Criterion 2 fails, some computational power is wasted on identifying elements in $\text{in}(A)$, $\text{out}(A)$, and $\text{OUT}(A)$. This can be relatively large, especially for large FoDs and cores. However, $|\mathfrak{F}_\Theta|/(2^{|\Theta|} - 1) \ll 1$ in most real-world applications, and under these conditions, the CCT method offers significant computational gains. Thus, one may even set the evaluation criteria based on $|\mathfrak{F}_\Theta|$ and $|A|$ only.

8.4 Experiments

Here we carry out several experiments to study the performance of the CCT method. The results demonstrate how the cost associated with conditional mass computations is related to certain properties of the BoE being conditioned and the conditioning event itself. Full knowledge of these relationships gives valuable information on how to optimize the conditional mass computation for efficient implementation of evidence update and fusion strategies [52].

8.4.1 Data Generation

For simulation and verification purposes, for a selected pair of values for $|\Theta|$ and $|\mathfrak{F}_\Theta| \in \{1, \dots, 2^{|\Theta|}\}$, we randomly generate a BoE $\mathcal{E} = \{\Theta, \mathfrak{F}_\Theta, m_\Theta\}$. Even for modest values of $|\Theta|$, the number of propositions possessing a positive belief (i.e., $|\hat{\mathfrak{F}}_\Theta|$) can be extremely high. Therefore, instead of computing the conditional masses for *all* the propositions in $\hat{\mathfrak{F}}_\Theta$, we select conditioning propositions $A_{\ell,j}$ as follows: select a value for \mathfrak{N} from $\{1, \dots, 2^{|\Theta|}\}$, and then randomly pick at most \mathfrak{N} conditioning propositions $A \in \hat{\mathfrak{F}}_\Theta$ for which $|A| = \ell$; repeat this for each value of ℓ in $\{1, \dots, |\Theta|\}$ so that we

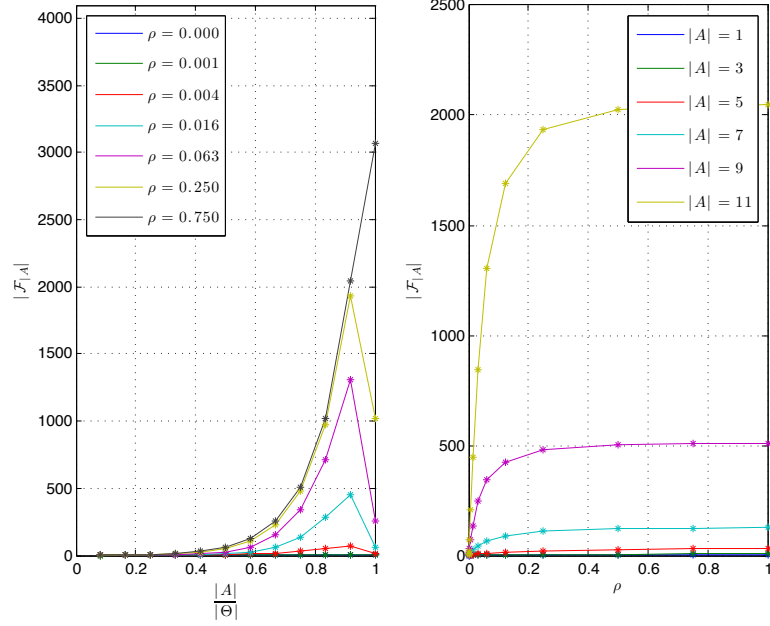


Figure 8.2: Variation of $|\mathfrak{F}_{\Theta|A}|$ versus $|A|/|\Theta|$ and ρ , where $|\Theta| = 12$.

end up with conditioning propositions of all cardinalities, but not more than \mathfrak{N} of each cardinality.

We repeat this procedure K times thus generating K different BoEs for each $\{|\Theta|, |\mathfrak{F}_{\Theta}|\}$ pair. The reported results are obtained by averaging over these K different BoEs; we used $K = 50$ in our experiments. We also use $\rho = |\mathfrak{F}_{\Theta}|/(2^{|\Theta|} - 1)$ as a measure of ‘density’ of the BoE’s core.

All simulations were carried out on an Apple Mac Pro Desktop with $2 \times 2.66\text{GHz}$ quad-core Intel Xeon processors, 16GB 1066MHz DDR3 RAM, running Mac OS X 10.6.2. Programs are non-optimized and no multiprocessing is used. Computational times were obtained while running regular programs. For each setting of parameters, $m_{\Theta}(\cdot|A)$ is computed using both the CCT and the standard methods.

8.4.2 Results and Analysis

8.4.2.1 Cardinality of the Conditional Core

For a given BoE, the cardinality of the conditional core $|\mathfrak{F}_{\Theta|A}|$ depends heavily on ρ and $|A|$. The left-hand side plot in Fig. 8.2 shows this variation for different ρ values. As can be seen, for $|A| < |\Theta|$, $|\mathfrak{F}_{\Theta|A}|$ increases exponentially with $|A|$; however, $|\mathfrak{F}_{\Theta|A}|$ sees a sudden drop at $|A| = |\Theta|$. The CCT explains this phenomenon well. As $|A|$ increases, the number of elements in $\text{out}(A)$ increases thus exponentially increasing $|\text{OUT}(A)|$ and $|\mathfrak{F}_{\Theta|A}|$. However, as $|A| = |\Theta|$, the number of elements in $\text{out}(A)$ suddenly decreases thus resulting in the sudden drop of $|\mathfrak{F}_{\Theta|A}|$.

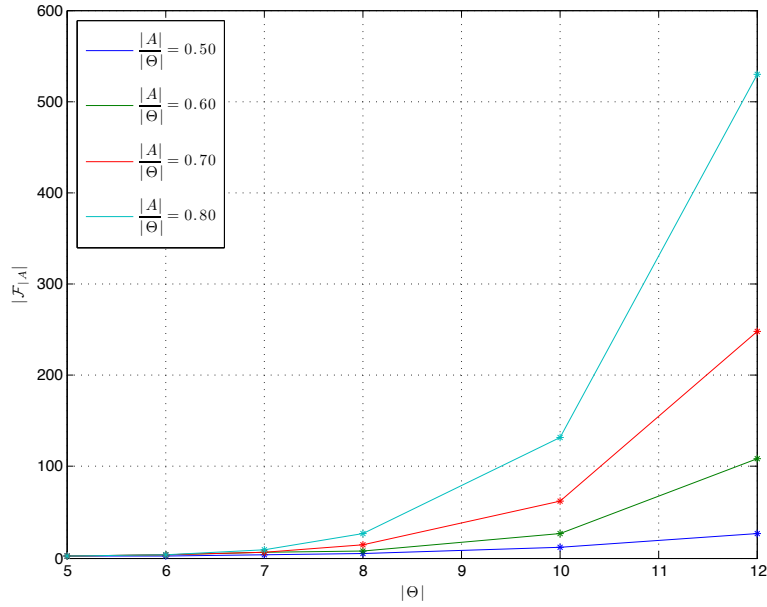


Figure 8.3: Variation of $|\mathfrak{F}_{\Theta|A}|$ versus $|\Theta|$ for different $|A|/|\Theta|$, with $\rho = 1/16$.

The right-hand side plot in Fig. 8.2 shows the variation of $|\mathfrak{F}_{\Theta|A}|$ versus ρ for different values of $|A|$. Clearly, $|A|$ imposes an upper bound on $|\mathfrak{F}_{\Theta|A}|$ because $|\mathfrak{F}_{\Theta|A}| \leq 2^{|A|} - 1$. When ρ is small (i.e., when there are only a very few focal elements present), $|\mathfrak{F}_{\Theta|A}|$ is determined by ρ ; but as ρ increases, the impact of ρ on $|\mathfrak{F}_{\Theta|A}|$ diminishes.

This scenario can also be understood via the CCT. Even though $|\mathfrak{F}_\Theta|$ and hence ρ increase, the $|\text{out}(A)|$ is always bounded by $|A|$. Therefore, an increasing ρ does not necessarily increase $|\text{OUT}(A)|$. Note that, the operational area of many applications, especially those dealing with soft data, correspond to FoDs with high cardinality with only a few number of (mostly non-singleton) propositions receiving positive support. This situation is depicted in the bottom left-hand corner of the plot where ρ is low and $|\mathfrak{F}_{\Theta|A}|$ is highly dynamic.

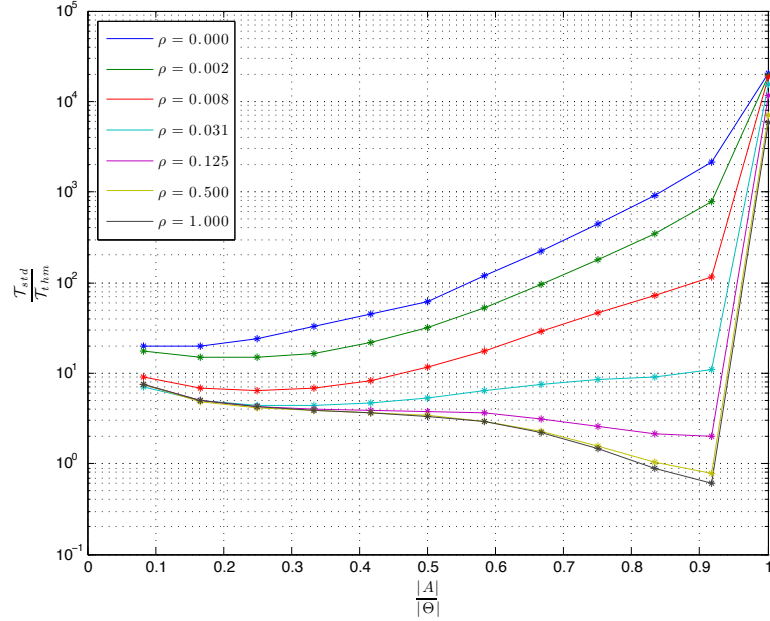


Figure 8.4: $\mathcal{T}_{std}/\mathcal{T}_{thm}$ versus $|A|/|\Theta|$ for different ρ , with $|\Theta| = 12$.

Fig. 8.3 shows the variation of $|\mathfrak{F}_{\Theta|A}|$ versus $|\Theta|$ for different values of $|A|/|\Theta|$, with fixed $\rho = 1/16$. Note how $|\mathfrak{F}_{\Theta|A}|$ increases exponentially with increasing $|\Theta|$.

8.4.2.2 Computational Time

Even though the absolute times are highly platform- and system- dependent, the ratio $\mathcal{T}_{std}/\mathcal{T}_{thm}$ provides a good measure of the relative speed. Fig. 8.4 shows the

variation of $\mathcal{T}_{std}/\mathcal{T}_{thm}$ (log scale) versus $|A|$ for different values of ρ , with fixed $|\Theta| = 12$.

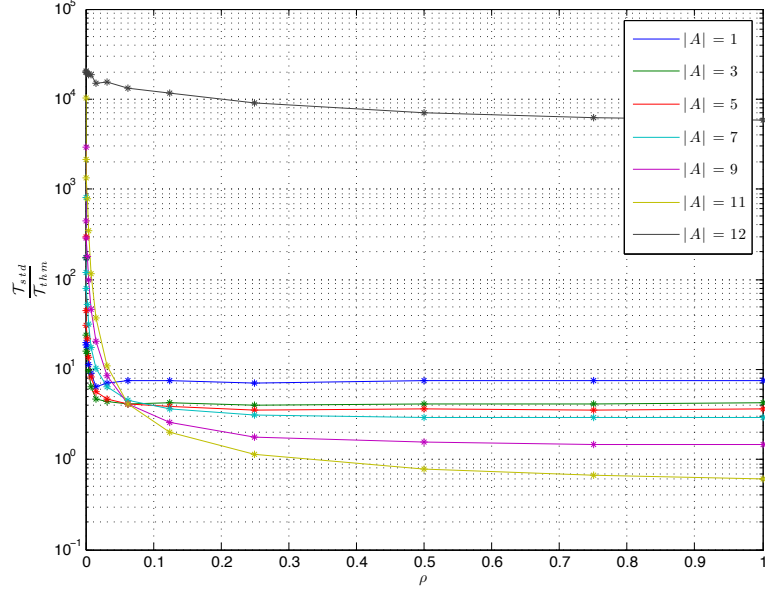


Figure 8.5: $\mathcal{T}_{std}/\mathcal{T}_{thm}$ versus ρ for different $|A|$, with $|\Theta| = 12$.

For approximately $\rho < 0.125$, all curves follow the same trend, viz., $\mathcal{T}_{std}/\mathcal{T}_{thm}$ increases with increasing $|A|$. Also, except when $\rho \geq 0.5$, $\mathcal{T}_{std}/\mathcal{T}_{thm} > 1$. Note that, with $|\Theta| = 12$, a $\rho \geq 0.5$ implies a BoE with over 2047 focal elements, which is rather unlikely in most practical applications; with $\rho = 0.02$, we still get a BoE with over 80 focal elements.

Fig. 8.5 shows the variation of $\mathcal{T}_{std}/\mathcal{T}_{thm}$ (log scale) versus ρ for different values of $|A|$, with fixed $|\Theta| = 12$. Note that, except when $|A| = 11$, $\mathcal{T}_{std}/\mathcal{T}_{thm} > 1$ for all values of ρ and $|A|$. Also, $\mathcal{T}_{std}/\mathcal{T}_{thm}$ is significantly large for smaller ρ , which is typical of many practical applications. Fig. 8.6 shows a zoomed-in version of Fig. 8.5 for smaller ρ , where $\mathcal{T}_{std}/\mathcal{T}_{thm}$ is seen to be quite high.

As Figs 8.4 and 8.5 show, the CCT method may not yield computational gains in all situations. The plot of $\mathcal{T}_{std} - \mathcal{T}_{thm}$ versus ρ in Fig. 8.7 perhaps illustrates this

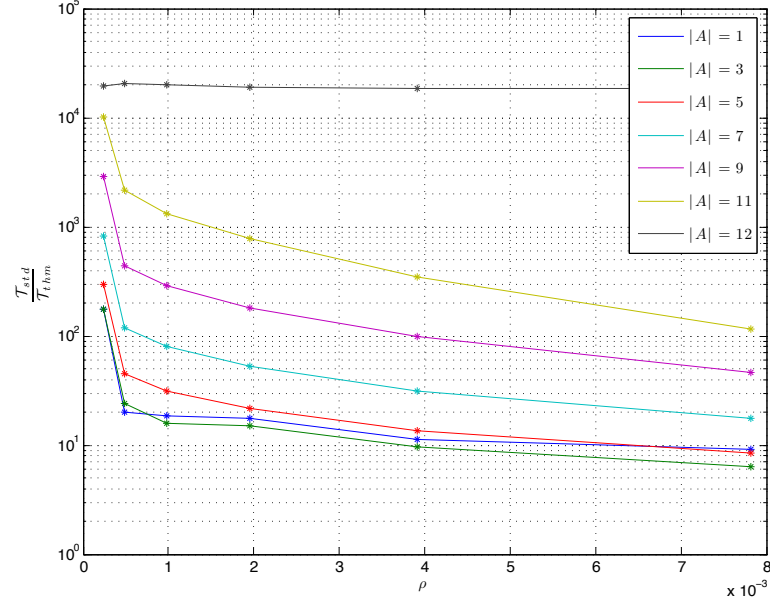


Figure 8.6: $\mathcal{T}_{std}/\mathcal{T}_{thm}$ versus smaller ρ for different $|A|$, with $|\Theta| = 12$.

better. Take, for example, the $|A| = 11$ case. It clearly shows that the CCT method is more expensive than the standard method for approximately $\rho > 0.3$. In this scenario, the additional overhead associated with identifying and processing of conditional focal elements becomes more expensive than a brute force direct evaluation of the conditional masses.

In most practical applications, the mass assignment strategy generates a very small number of focal elements. The problem with the DS belief computation is that, even though $|\mathfrak{F}_\Theta|$ may remain significantly small, the computational time increases exponentially as $|\Theta|$ increases. Fig. 8.8 compares the variations of \mathcal{T}_{std} and \mathcal{T}_{thm} versus $|\Theta|$ for different values of $|A|$, with fixed $|\mathfrak{F}_\Theta| = 4$. It clearly shows the exponential increase of the computational time \mathcal{T}_{std} associated with the standard method. In contrast, the time \mathcal{T}_{thm} associated with the CCT method remains almost constant, thus overcoming the scaling issue of the DS theoretic conditional mass computation to a large extent.

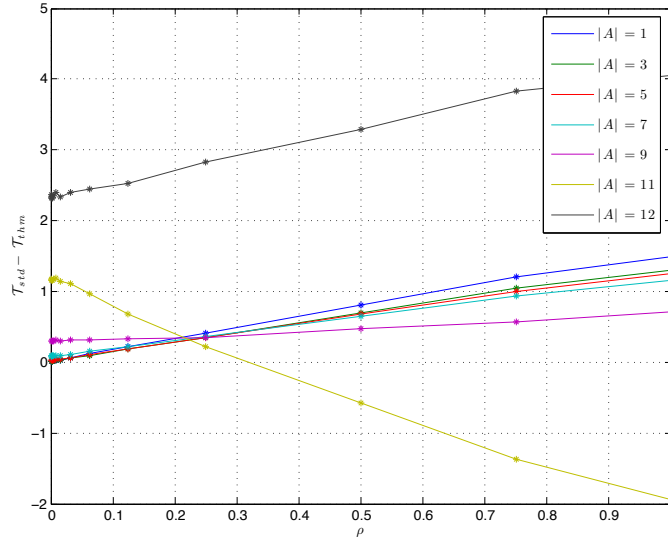


Figure 8.7: $\mathcal{T}_{std} - \mathcal{T}_{thm}$ versus ρ for different $|A|$, with $|\Theta| = 12$.

8.5 Chapter Summary

The proposed graphical method along with an efficient message-passing scheme based on the CCT, clearly reduces the number of conditional mass computations. But, the total computational cost, as we have pointed out, depends on the implementation and the scenario under consideration. For larger FoDs, the proposed strategy achieves very high computational gains. Moreover, with the cardinalities of the conditioning event and the core kept fixed, the computational cost of CCT method remains almost constant with increasing cardinality of the FoD. Thus, the CCT method scales well for large FoDs. The standard method becomes prohibitive in such a scenario. We believe that this constitutes a very significant result, especially when dealing with soft/hard data fusion scenarios where one has to deal with rather large FoDs possessing relatively smaller number of focal elements. Of course, the CCT method may not provide computational gains in all operating conditions. The proposed empirical criteria can be used to identify the conditions that would render the use of the CCT

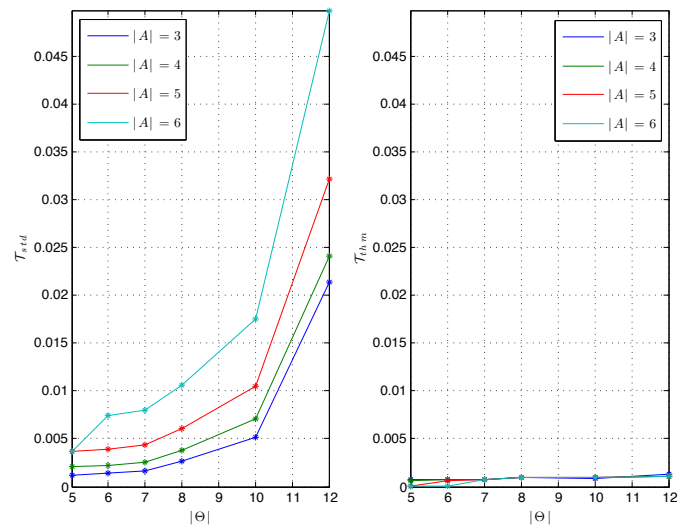


Figure 8.8: \mathcal{T}_{std} and \mathcal{T}_{thm} versus $|\Theta|$, for different $|A|$, with $|\mathfrak{F}_\Theta| = 4$.

method computationally more efficient.

DST Core Approximations

As we have pointed out earlier, the computational overhead of DST methods are relatively much higher than, for instance, Bayesian methods mainly due to the fact that DST methods work with the powerset of the source FoDs. On the other hand, this very fact is the main reason for the enormous flexibility that the DS theory provides in uncertain data modeling. In Chapter 8, we illustrated how the CCT can be exploited to directly identify the conditional core thus potentially providing significant computational gains. In this Chapter, we propose yet another strategy to reduce the computational burden associated with DST methods: approximation of DST data models based on statistical sampling techniques. Our proposed method allows one to approximate the focal set based on an objective function and then statistically redistribute the masses of removed focal elements. The objective function can be chosen depending on the application to remove irrelevant propositions (e.g., those that are impossible to occur). In the absence of such information, the proposed method can merely impose bounds on the cardinality or the minimum mass of focal elements (similar to the methods in [97–100]).

This chapter is organized as follows: Section 9.1 provides a chapter overview; Section 9.2 provides a review of existing DST approximation methods; Section 9.3

presents the main result of this chapter; Section 9.4 contains an extensive set of simulations highlighting the salient points of the presented method; and finally Section 9.5 provides the chapter summary.

9.1 Overview

As we have already discussed, potential of many DST methods is often thwarted by the significant computational overhead in terms of storage and processing. Computational optimization techniques (e.g., fast computational methods and approximations) to address this issue have been proposed in the DST literature. For example, the extensive amount of work carried out by Wilson, et al., [96, 101–103] addresses issues related to fast belief/plausibility computation, DCR approximations, and approximated decision making.

In real-life applications, evidence can usually be captured via a smaller number of focal elements which is far less than the possible maximum (i.e., powerset of the FoD). However, repeated DST operations (e.g., evidence combination, conditioning) have the potential to exponentially increase the number of focal elements, thus making the subsequent processing more expensive. Given the fact that storage requirements and processing costs are directly proportional to the number of focal elements, a natural question that arises is whether one can reduce the computational overhead in terms of storage and processing by explicitly reducing the number of focal elements through approximations.

The approach taken by most of the existing techniques is to retain only the propositions with highest masses and recompute their masses via redistribution and/or normalization [97–100]. Even though, some of these methods are as simple as summarizing the mass of discarded propositions to a single proposition obtained as the disjunction of the discarded propositions, their performance is very satisfactory in most practical applications. However, one must keep in mind that removal of focal

elements and redistribution of their masses have to be done with extreme care to preserve the underlying evidence and the associated uncertainties to a level that does not alter the final inferences. Therefore, methods that simply retain the highest valued propositions may not be appropriate for all applications and may not be able to capture the underlying meaning adequately well.

Approximation approaches such as statistical sampling methods have gained importance in many probabilistic signal processing methods. For example, sampling methods are commonly used in optimal filtering and tracking (e.g., in particle filters), where complicated posterior probabilities are approximated in a formal manner. Several approximation techniques based on statistical sampling methods have been used in DST as well, e.g., approximating the combined belief [101]. However, to the best of our knowledge, such sampling methods have not been used for approximation of focal elements and their support.

9.2 Approximations of the Core

In this section, we provide a review of some of the existing approximation techniques that reduce the computational burden by reducing the number of focal elements. Suppose the BoE $\mathcal{E}_\Theta = \{\Theta, \mathfrak{F}_\Theta, m_\Theta(\cdot)\}$ is to be approximated by the new BoE $\mathcal{E}'_\Theta = \{\Theta, \mathfrak{F}'_\Theta, m'_\Theta(\cdot)\}$.

9.2.1 The Bayesian Approximation (BA)

This method reduces a given BPA to a PMF [97]. Thus, only singleton propositions are allowed in the approximated BPA.

Definition 38 (Bayesian Approximation (BA)). [97] *The BoE \mathcal{E}'_Θ is given by*

$$m'_\Theta(B) = \begin{cases} \frac{1}{K_{BA}} \sum_{B \subseteq C} m_\Theta(C), & \text{for } |B| = 1; \\ 0, & \text{otherwise.} \end{cases}$$

where $K_{BA} = \sum_{C \subseteq \Theta} m(C) |C|$. ■

By definition, $|\mathfrak{F}'_\Theta|$ is at most $|\Theta|$ and the cost of approximation is in the order of $\mathcal{O}(|\mathfrak{F}_\Theta| \cdot |\Theta|)$. Furthermore, if the BoEs are combined using the DCR, then the combination and approximation do not depend on the order, i.e., one can either combine BoEs prior to the approximation, or vice versa, and obtain the same result. The BA is the only approximation method that possesses this property.

9.2.2 The k - ℓ - x Method ($k\ell x$)

This method focuses on retaining only the highest valued focal elements [98]. The BoE \mathcal{E}'_Θ will have at least k or at most ℓ focal elements with a sum of the BPA being at least $1 - x$, $x \in [0, 1]$. The approximation is finally normalized s.t. $\sum_{B \in \mathfrak{F}'_\Theta} m'_\Theta(B) = 1$. The approximation time is in the order of $\mathcal{O}(|\mathfrak{F}_\Theta| \cdot \log(|\mathfrak{F}_\Theta|))$.

9.2.3 Summarization Method (SM)

This method leaves the highest valued $k - 1$ focal elements intact and ‘summarizes’ the remaining focal elements to their set theoretic union [99].

Definition 39 (Summarization Method (SM)). [99] *Let k be the number of focal elements to be contained in \mathcal{E}'_Θ and let \mathcal{M} denote the set of $k - 1$ focal elements*

$B \in \mathfrak{F}_\Theta$ with the highest BPA. Then, the BoE \mathcal{E}'_Θ is given by

$$m'_\Theta(B) = \begin{cases} m_\Theta(B), & \text{for } B \in \mathcal{M}; \\ \sum_{C \in \mathfrak{F}_\Theta \setminus \mathcal{M}} m_\Theta(C), & \text{for } A = A_0; \\ 0, & \text{otherwise,} \end{cases}$$

where $A_0 = \bigcup_{C \in \mathfrak{F}_\Theta \setminus \mathcal{M}} C$. ■

This approximation is extremely fast and can be computed in $\mathcal{O}(|\mathfrak{F}_\Theta|)$, even though the applicability to arbitrary BoEs is arguable.

9.2.4 D1 Approximation (D1)

This method retains a set of highest valued focal elements and distributes the BPA of the remaining focal elements among them [100]. The BPA distribution is intuitive and the method is applicable to arbitrary BoEs.

Let k be the desired number of focal elements to be contained in \mathcal{E}'_Θ and let \mathcal{M}^+ denote the set of $k - 1$ focal elements $B \in \mathfrak{F}_\Theta$ with the highest BPA, and $\mathcal{M}^- = \mathfrak{F}_\Theta \setminus \mathcal{M}^+$. Then, the BPA of the focal elements in $B \in \mathcal{M}^-$ is distributed among the elements in \mathcal{M}^+ as follows:

Given a $B \in \mathcal{M}^-$, compute $\mathcal{M}_B = \{C \in \mathcal{M}^+ \mid B \subseteq C\}$. Then $m_\Theta(B)$ is dispensed uniformly among the set-theoretically smallest members of \mathcal{M}_B . If $\mathcal{M}_B = \emptyset$, then $\mathcal{M}'_B = \{C \in \mathcal{M}^+ \mid |C| \geq |B|, C \cap B \neq \emptyset\}$ is generated and $m_\Theta(B)$ is shared among the smallest members. This process is invoked recursively until all of $m_\Theta(B)$ is assigned to \mathcal{M}^+ or \mathcal{M}'_B is empty, in which case the remaining mass is assigned to Θ . The cost of the D1 approximation is $\mathcal{O}(k \cdot (|\mathfrak{F}_\Theta| - k))$. This approximation is conservative in the sense that the BoE \mathcal{E}'_Θ is less specific than the original \mathcal{E}_Θ .

9.3 Monte Carlo Core Approximation (MCCA)

In this section, we introduce our *Monte Carlo core approximation (MCCA) technique*, a sampling-based technique for approximating the core for the purpose of computational overhead reduction in DST methods. Monte Carlo (MC) approach is commonly used in statistical signal processing literature to estimate complex analytic or unknown probability distributions with sample-based representations. Here, we use the MC method to estimate a given BoE \mathcal{E}_Θ with a new BoE \mathcal{E}'_Θ such that **(a)** \mathcal{E}'_Θ is computationally more efficient than \mathcal{E}_Θ ; and **(b)** decisions generated with \mathcal{E}'_Θ are close to those generated with \mathcal{E}_Θ .

Regarding **(a)**, the computational gains are to be achieved by reducing the number of focal elements in the core via an objective function that chooses the focal elements to be retained in \mathcal{E}'_Θ . This objective function can be chosen to obtain an ‘optimal’ core depending on the application. In a simple setup, it can be chosen to limit the number of focal elements (e.g., pick the focal elements with the highest BPA). However, it can also be used to satisfy more elaborate properties (e.g., to avoid certain composite focal elements with impossible singleton combinations).

Regarding **(b)**, the probability mass function (PMF) $BetP(\cdot)$ (generated via the pignistic transformation) is often used in DST for decision-making. Thus, the new BoE \mathcal{E}'_Θ is generated s.t. the pignistic transformation $BetP(\cdot)$ corresponding to $m_\Theta(\cdot)$ given is approximately equal to the pignistic transformation $BetP'(\cdot)$ corresponding to $m'_\Theta(\cdot)$. Let us proceed by formally stating the approximation problem.

9.3.1 Problem Formulation

Let $\mathcal{E}_\Theta = \{\Theta, \mathfrak{F}_\Theta, m_\Theta(\cdot)\}$ be the BoE to be approximated by the BoE $\mathcal{E}'_\Theta = \{\Theta, \mathfrak{F}'_\Theta, m'_\Theta(\cdot)\}$. The core \mathfrak{F}'_Θ is to be determined by an objective function $\mathfrak{D} : 2^\Theta \mapsto 2^\Theta$ with $\mathfrak{D}(\mathfrak{F}_\Theta) = \mathfrak{F}'_\Theta$ s.t. $|\mathfrak{F}'_\Theta| < |\mathfrak{F}_\Theta|$. The new BPA $m'_\Theta(\cdot) : 2^\Theta \mapsto [0, 1]$ is to be derived s.t. $BetP'(\cdot) \approx BetP(\cdot)$, where $BetP'(\cdot)$ and $BetP(\cdot)$ are the pignistic

transformations of $m'_\Theta(\cdot)$ and $m_\Theta(\cdot)$, respectively.

9.3.2 Algorithm

Once the \mathfrak{F}'_Θ is generated via the objective function $\mathfrak{D}(\cdot)$, we use the MC method for approximating $m'_\Theta(\cdot)$ s.t. $BetP'(\cdot) \approx BetP(\cdot)$. The approximation procedure is as follows:

Step I. Initialization:

1. Define the collections \mathcal{G}_θ as

$$\mathcal{G}_\theta \equiv \{B \in \mathfrak{F}_\Theta \mid \theta \subseteq B\}.$$

2. Define the corresponding PMF $P_{\mathcal{G}}(B \mid \theta) : \mathcal{G}_\theta \mapsto [0, 1]$ as

$$P_{\mathcal{G}}(B \mid \theta) = \frac{1}{\sum_{C \in \mathcal{G}_\theta} m_\Theta(C)} m_\Theta(B).$$

3. Initialize the weights $\mathcal{W}_B(0)$ as

$$\mathcal{W}_B(0) = \begin{cases} m_\Theta(B), & \text{for } B \in \mathfrak{F}'_\Theta; \\ 0, & \text{otherwise.} \end{cases}$$

4. Define the weight distribution constant $\mathcal{K}_{\mathfrak{F}_\Theta \setminus \mathfrak{F}'_\Theta}$ as

$$\mathcal{K}_{\mathfrak{F}_\Theta \setminus \mathfrak{F}'_\Theta} = 1 - \sum_{C \in \mathfrak{F}'_\Theta} m_\Theta(C).$$

Step II. Sampling:

1. Sample $\theta_k \in \Theta$, $k = 1, \dots, N_s$, from $BetP(\theta)$.
2. Sample $B_k \in \mathcal{G}_{\theta_k}$ from $P_{\mathcal{G}}(B \mid \theta_k)$.
3. If $B_k \in \mathfrak{F}'_\Theta$, update the weight of B_k as

$$\mathcal{W}_{B_k}(k) = \mathcal{W}_{B_k}(k-1) + \frac{1}{N_s} \mathcal{K}_{\mathfrak{F}_\Theta \setminus \mathfrak{F}'_\Theta}.$$

4. If $B_k \notin \mathfrak{F}'_\Theta$, then resample.

Step III. Resampling:

1. Define \hat{G}_{B_k} and the corresponding PMF $\hat{P}_G(B \mid B_k)$ as follows:

- (a) Let $\hat{G}_{B_k} = \{B \in \mathfrak{F}'_\Theta \mid B_k \subseteq B\}$.

- (b) If $\hat{G}_{B_k} \neq \emptyset$, then do the following:

- i. Define the PMF $\hat{P}_G(B \mid B_k)$ as

$$\hat{P}_G(B \mid B_k) = \frac{1}{\hat{\mathcal{L}}_{B_k}} m(B),$$

$$\text{where } \hat{\mathcal{L}}_{B_k} = \sum_{C \in \hat{G}_{B_k}} m(C).$$

- ii. Sample $\hat{B}_k \in \hat{G}_{B_k}$ from $\hat{P}_G(B \mid B_k)$.

- iii. Update the weight of B_k as

$$\mathcal{W}_{B_k}(k) = \mathcal{W}_{B_k}(k-1) + \frac{1}{N_s} \mathcal{K}_{\mathfrak{F}_\Theta \setminus \mathfrak{F}'_\Theta}.$$

- (c) If $\hat{G}_{B_k} = \emptyset$, then do the following:

- i. Redefine \hat{G}_{B_k} as

$$\hat{G}_{B_k} = \{B \in \mathfrak{F}'_\Theta \mid B \cap B_k \neq \emptyset\}.$$

- ii. If $\hat{G}_{B_k} \neq \emptyset$, update the weights of $B \in \hat{G}_{B_k}$ as

$$\mathcal{W}_B(k) = \mathcal{W}_B(k-1) + \frac{1}{N_s \mathcal{L}_{B_k}^{\sup}} \frac{|B \cap B_k|}{|B \cup B_k|} \mathcal{K}_{\mathfrak{F}_\Theta \setminus \mathfrak{F}'_\Theta},$$

where

$$\mathcal{L}_{B_k}^{\sup} = \sum_{B \in \hat{G}_{B_k}} \frac{|B \cap B_k|}{|B \cup B_k|}.$$

iii. If $\hat{G}_{B_k} = \emptyset$, update the weights of $B \in \mathfrak{F}'_{\Theta}$ as

$$\mathcal{W}_B(k) = \mathcal{W}_B(k-1) + \frac{1}{N_s |\mathfrak{F}'_{\Theta}|} \mathcal{K}_{\mathfrak{F}_{\Theta} \setminus \mathfrak{F}'_{\Theta}}.$$

2. Select the BPA $m'_{\Theta}(\cdot)$ as

$$\begin{aligned} m'_{\Theta}(B) &= \sum_{C \subseteq \Theta} \mathcal{W}_C[N_s] \delta(B - C) \\ &= \begin{cases} \mathcal{W}_B[N_s], & \text{for } B \in \mathfrak{F}'_{\Theta}; \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

To compare the various approximation methods with the proposed MCCA method, let us consider the following example taken from [100].

Example 17. [100] The BoE \mathcal{E}_{Θ} for which $\Theta = \{a, b, c, d, e\}$ and $m_{\Theta}(\{ab, acd, c, cd, de\}) = \{0.50, 0.30, 0.10, 0.05, 0.05\}$ is to be approximated by \mathcal{E}'_{Θ} . The approximated results are tabulated in Table 9.1.

Approximation parameters for methods BA, $k\ell x$, SM and D1 are chosen as in [100]; for the MCCA, we choose $|\mathfrak{F}'_{\Theta}| = 3$ with $N_s = 100$. \square

Method	\mathfrak{F}'_{Θ}	m'_{Θ}
BA	$\{a, b, c, d, e\}$	$\{0.360, 0.230, 0.205, 0.180, 0.023\}$
$k\ell x$	$\{ab, acd, e\}$	$\{0.556, 0.333, 0.111\}$
SM	$\{ab, acd, cde\}$	$\{0.500, 0.300, 0.200\}$
D1	$\{ab, acd, \Theta\}$	$\{0.500, 0.475, 0.025\}$
MCCA	$\{ab, acd, c\}$	$\{0.538, 0.354, 0.108\}$

Table 9.1: Core Approximation of Example 17

9.3.3 Selection of an Appropriate Objective Function

After a several iterations of processing, the BoE may contain “unwanted” propositions (e.g., impossible combinations of singletons) that may not necessarily have the lowest support. Combination and conditioning operations often generate large numbers of propositions, which often grow exponentially with respect to the number of operations (see Section 9.4 for details).

In MCCA, via the objective function, one has the flexibility of specifying what propositions are to be kept and/or removed (without being forced to retain the propositions having the highest mass). The choice is of course application dependent. To illustrate this, let us take an example involving DST conditioning.

Example 18. Consider a BoE $\mathcal{E}_\Theta = \{\Theta, \mathfrak{F}_\Theta, m_\Theta(\cdot)\}$ with $\Theta = \{a, b, c, \dots, x, y, z\}$, $\mathfrak{F}_\Theta = \{b, a, pqr, bck, dl, em\}$ and $m_\Theta(B) = \{0.35, 0.25, 0.15, 0.10, 0.10, 0.05\}$, for $B \in \mathfrak{F}_\Theta$ (in the same order given in \mathfrak{F}_Θ). Let us compute the conditional masses with respect to the conditioning event $A = (abcdefghij)$ (see Table 9.2).

B	$m(B A)$	B	$m(B A)$	B	$m(B A)$	B	$m(B A)$
a	0.2941	$abcd$	0.0121	b	0.4118	bcd	0.0169
abc	0.0392	$abce$	0.0054	bc	0.0549	bce	0.0076
ad	0.0392	ade	0.0054	bd	0.0549	bde	0.0076
ae	0.0184	$abcde$	0.0029	be	0.0257	$bcde$	0.0039

Table 9.2: Conditional Core Corresponding to $A = (abcdefghij)$.

Notice the following. When the BoE is conditioned with respect to $A = (abcdefghij)$, all the focal elements that are not contained in A vanish. The elements that “straddle” A and \bar{A} (i.e., elements in $out(A)$) generate a slew of conditional focal elements by making arbitrary unions. Thus, conditioning increases the size of the core from 7

to 16, even though the “focus” has narrowed down to 10 ($= |A|$) elements from 20 ($= |\Theta|$).

Let us assume that, for the application at hand, we are only interested in retaining the focal elements that are “contained in” the conditioning event. However, to preserve the contribution from focal elements that straddle A and \bar{A} , we also retain the focal elements that are unions of **(i)** an element in $\text{in}(A)$ and **(ii)** the union of all the elements in $\text{out}(A)$ (i.e., largest (set theoretic) element in $\text{OUT}(A)$). With the help of CCT, one can define the objective function to achieve exactly this.

	Approximated BoE		Error		
	\mathcal{F}'	m'	$ERR1$	RMS	MAE
BA	a, b, c, d, e	.293, .452, .100, .100, .055	.057	.031	.214
$k\ell x$	b, a, bd, bc	.505, .361, .067, .067	.063	.026	.162
SM	$b, a, bd, abcde$.412, .294, .055, .239	.022	.010	.060
$D1$	b, a, bd, bc	.503, .387, .055, .055	.049	.027	.186
$MCCA^\dagger$	$b, a, bcde, abcde$.508, .365, .038, .089	.040	.021	.132
$MCCA^\ddagger$	b, a, bd, bc	.479, .354, .083, .084	.053	.021	.128

Table 9.3: Core Approximation of Example 18

Approximated results appear in Table 9.3. In Table 9.3, $MCCA^\dagger$ retains the focal elements of interest only. As a means of comparing with the BA , $k\ell x$, SM and $D1$ methods, $MCCA^\ddagger$ retains the focal elements with the highest mass only. As we can see, the results of $MCCA^\dagger$ is comparable to other methods, while it retains more meaningful propositions as required by the application. The error measure that are used in Table 9.3 are defined in Section 9.4.3. \square

9.4 Experiments

In this section, we carry out an experiment to analyze the behavior of the MCCA as well as to compare its performance to existing methods. Full knowledge of the behavior and comparative results to exiting algorithms provide valuable information on how to optimize and also to choose an appropriate algorithm for a given application. We proceed by explaining the experimental methodology.

9.4.1 Methodology

Consistent with what has been employed in [98], and later followed in [100], the methodology we adopted for constructing the BoEs is the following:

1. Set $|\Theta| = 32$ and use an exponential distribution [100] to generate 6 random BoEs $\mathcal{E}_{\Theta_i} \equiv \{\Theta, \mathfrak{F}_{\Theta_i}, m_{\Theta_i}(\cdot)\}$, $i = 1, \dots, 6$, s.t. each focal set has 8 elements, i.e., $|\mathfrak{F}_{\Theta_i}| = 8$, $i = 1, \dots, 6$.
2. Using the DCR, fuse the BoEs as

$$\mathcal{E}_{\Theta}[j] = \begin{cases} \mathcal{E}_{\Theta_1} \oplus \mathcal{E}_{\Theta_2}, & \text{for } j = 1; \\ \mathcal{E}_{\Theta}[j-1] \oplus \mathcal{E}_{\Theta_{j+1}}, & \text{for } j = 2, \dots, 5. \end{cases} \quad (9.1)$$

Let $\mathcal{E}_{\Theta} = \mathcal{E}_{\Theta}[5]$.

3. Generate the approximations

$$\mathcal{E}_{\Theta_X}[j] = \begin{cases} X(\mathcal{E}_{\Theta}[1]), & \text{for } j = 1; \\ X(\mathcal{E}_{\Theta_X}[j-1] \oplus \mathcal{E}_{\Theta_{j+1}}), & \text{for } j = 2, \dots, 5, \end{cases} \quad (9.2)$$

where $X(\mathcal{E}_{\Theta}[\cdot])$ denotes the approximation of $\mathcal{E}_{\Theta}[\cdot]$ with the method X . Let $\mathcal{E}'_{\Theta} = \mathcal{E}_{\Theta_X}[5]$.

4. Repeat the above procedure for all approximation methods BA, *klx*, SM, D1 and MCCA.

5. Repeat the whole procedure for 1000 times.

9.4.2 Parameters

For consistency, we choose the same approximation parameters used in [100]. But in contrast to [100], we use the same set of BoEs for all approximation methods. Even though the differences diminish as the number of iterations increase, this procedure allows for a fairer comparison.

We identify the various approximations as follows. The $k\ell x$ method carried out with $\{k = 7, \ell = 8, x = 0.0001\}$ and $\{k = 29, \ell = 30, x = 0.0001\}$ are denoted by $k\ell x-8$ and $k\ell x-30$, respectively. The experiments with the $k\ell x$ method in [100] are carried out in a way that one setting selects as many focal elements as it needs and another setting selects only one focal element. However, our parameter selection allows a fairer comparison by selecting the same number of focal elements for all the methods. The SM, D1, and MCCA methods with a maximum of 8 focal elements are denoted by SM-8, D1-8, and MCCA-8, respectively; with a maximum of 30 focal elements, the approximations are denoted by SM-30, D1-30, and MCCA-30, respectively.

9.4.3 Performance Criteria

We use the measures used in [98, 100] along with other frequently used error measures for comparison purposes.

9.4.3.1 Quantitative Measures

$$Error1 = \max_{\theta \in \Theta} |BetP(\theta) - BetP'(\theta)|; \quad (9.3)$$

$$RMS = \sqrt{\sum_{\theta \in \Theta} \frac{(BetP(\theta) - BetP'(\theta))^2}{|\Theta|}}; \quad (9.4)$$

$$MAE = \sum_{\theta \in \Theta} |BetP(\theta) - BetP'(\theta)|. \quad (9.5)$$

The measure *Error1* is used in [98], and later in [100].

9.4.3.2 Qualitative Measures

Let $\theta_0, \theta'_0 \in \Theta$ be the best choices among all the alternatives generated via $BetP(\bullet)$ and $BetP'(\bullet)$, respectively, i.e.,

$$\theta_0 = \arg \max_{\theta \in \Theta} BetP(\theta); \quad \theta'_0 = \arg \max_{\theta \in \Theta} BetP'(\theta).$$

Then, measures *Error2* and *Error3* are given by

$$\begin{aligned} Error2 &= |\{\theta \in \Theta \mid BetP'(\theta) > BetP'(\theta_0)\}|; \\ Error3 &= |\{\theta \in \Theta \mid BetP(\theta) > BetP(\theta'_0)\}|. \end{aligned} \quad (9.6)$$

The measures *Error2* and *Error3* are proposed and used in [100]. The measure *Error3* is particularly important for assessing an approximation method with respect to decision-making; *Error3* = 0 represents the case when the approximated BoE yields the same decision as that of the original BoE.

9.4.4 Results and Analysis

9.4.4.1 Behavior of the MCCA

To study the behavior of the MCCA method with respect to its parameters, we compare the approximation error of the MCCA-generated DCR-fused BoE \mathcal{E}'_{Θ} in (9.2) with the original DCR-fused BoE \mathcal{E}_{Θ} in (9.1) for different parameter configurations. The idea here is to understand the sensitivity of MCCA to its parameters, e.g., the number of samples N_s .

Variation of *Error1*. As one would expect, *Error1* decreases (see Fig. 9.1) with increasing cardinality which is to be expected because, as the size of the approximated core approaches the size of the original core, the approximation error

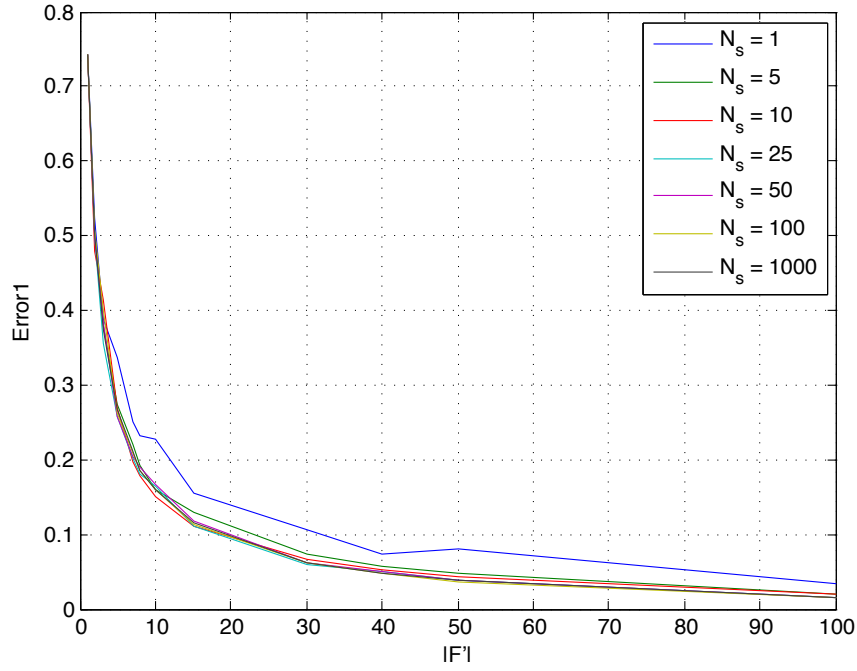


Figure 9.1: MCCA-generated approximation: $Error1$ Vs. $|\mathfrak{F}'_{\Theta}|$.

diminishes. However, it is important to notice that the error is not highly dependent on the number of sampling iterations (see Fig. 9.2). This can be understood since the approximation method attempts to preserve the underlying PMF obtained via the pignistic transformation. Behavior of other performance measures are also not highly dependent on the number of sampling iterations. Hence, we only show the dependency of performance measures on the cardinality of the approximated core.

Variations of $Error2$ and $Error3$. See Fig. 9.3. The variation of $Error2$ is somewhat arbitrary and small compared to $Error3$ which decreases as $|\mathfrak{F}'_{\Theta}|$ decreases. Again, the dependency of $Error2$ and $Error3$ on N_s is minimal.

Variations of RMS and MAE . See Fig. 9.4. This behavior is similar to that of $Error1$.

In summary, it can be seen that the proposed MCCA algorithm ‘converges’ in the sense that it is possible to pick the approximation parameters (e.g., N_s) to achieve a

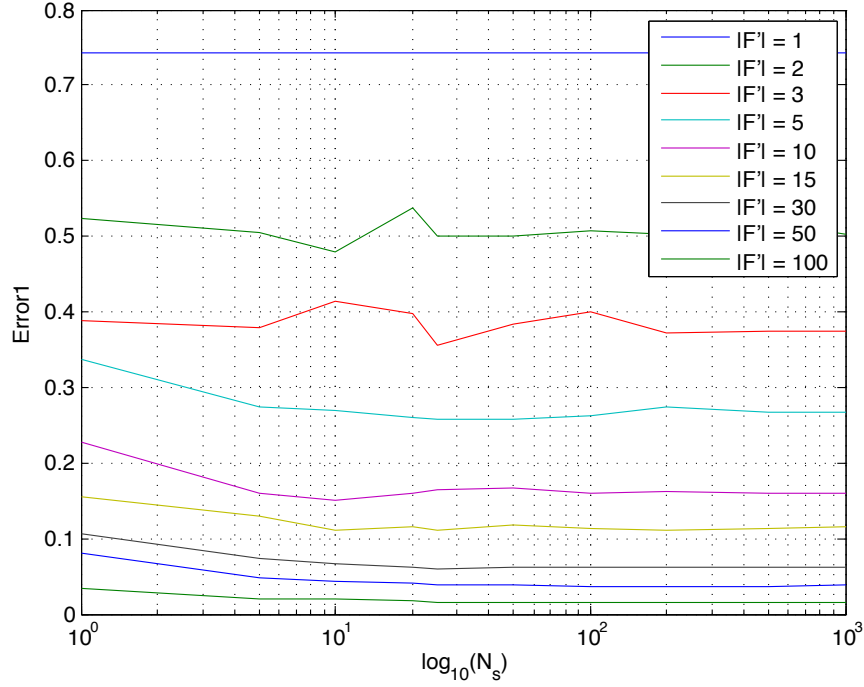


Figure 9.2: MCCA-generated approximation: $Error1$ Vs. N_s .

balance between a desired level of approximation error and the computational overhead.

9.4.4.2 Performance Comparison

In this section, we compare and contrast the error performance of MCCA to the other existing approximation techniques that were discussed in Section ??.

Cardinality of the Core. Table 9.4 shows the average, minimum, and maximum cardinality of the core generated at each step when the 6 BoEs are fused using the DCR to generate the final fused original BoE $\mathcal{E}_\Theta[k]$ in (9.1). The approximations in (9.2) generate either 8 or 30 focal elements at each step.

Variation of $Error1$. Fig. 9.5 shows the variation of $Error1$ of the i -th DCR-fused combination in (9.2). 30 focal elements cases of all methods outperform their

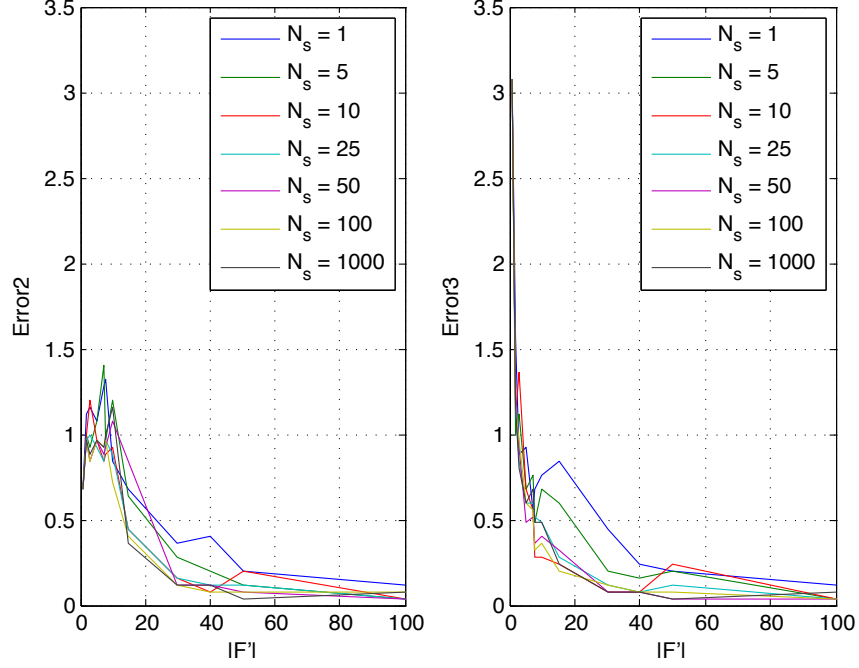


Figure 9.3: MCCA-generated approximation: $Error2$ and $Error3$ Vs. $|\mathfrak{F}'_{\Theta}|$.

8 focal elements counterparts as expected. It is important to note that the approximation techniques approximate an increasing number of focal elements. $MCCA - 30$ and $D1 - 30$ are superior to other methods. Error in BA method tends to decrease over the others. However, the use of BA approximation is very limited since it destroys the uncertainty associated with focal elements by only estimating masses for singletons.

Variations of $Error2$ and $Error3$. Fig. 9.6 shows the variations of $Error2$ and $Error3$ of the i -th DCR-fused combination in (9.2). $MCCA-30$ is among the other best alternatives in both measures. In fact, the error is almost zero everywhere, implying that the approximated BoE generates the same optimal choice as the original BoE. It is important to notice that $MCCA-8$ belongs to the group of 30 focal element approximations.

Variations of RMS and MAE . The results corresponding to the variations

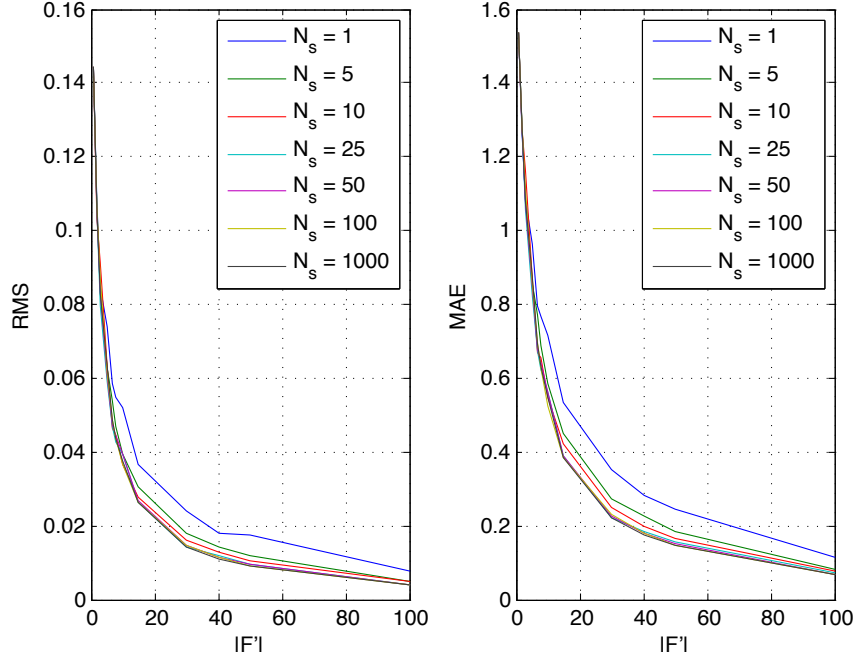


Figure 9.4: MCCA-generated approximation: RMS and MAE Vs. $|\mathfrak{F}'_{\Theta}|$.

of RMS and MAE measures (which provide aggregated overall errors) in Fig. 9.7 demonstrate features that are similar to observations made in Fig. 9.6.

Notice that the variations of the error measures in our experiments for the D1, SM, are consistent with the results in [100].

9.5 Chapter Summary

When compared to using measures that have been used in the literature, the proposed approximation technique is comparable to the existing methods in both qualitative and quantitative aspects. We have shown that the proposed methodology is robust and less sensitive to approximation parameters and provides accurate predictions with relatively smaller number of sampling iterations. However, some of the other approximation methods are simpler and more efficient to implement. Thus, one

Fusion Step	1	2	3	4	5
Average	63.9	433.2	1020.8	1027.3	782.9
Minimum	62	325	481	386	264
Maximum	64	494	1752	2329	1972

Table 9.4: Cardinality of the Core of the DCR-Fused BoE $\mathcal{E}_\Theta[k]$

needs to select an appropriate approximation strategy depending on the application. The results of the empirical study carried out in the experimental section can be used to aid this task.

One key feature of the proposed methodology is the flexibility of specifying the desired core, thus allowing for a more meaningful approximation than merely restricting to propositions with the highest support. In the case of conditioning, one can make use of the CCT to specify a more specific (in contrast to less specific) core for approximation via the MCCA as we have illustrated in an example.

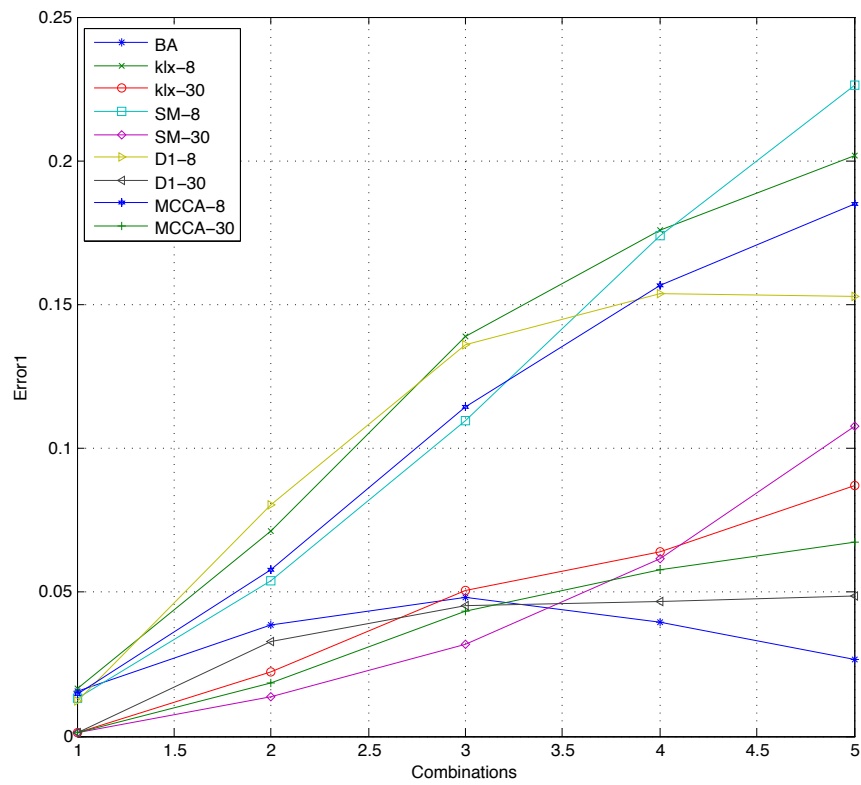


Figure 9.5: $Error1$ of $\mathcal{E}_{\Theta_X}[i]$, $i = 1, \dots, 5$, the i -th DCR-fused BoE.

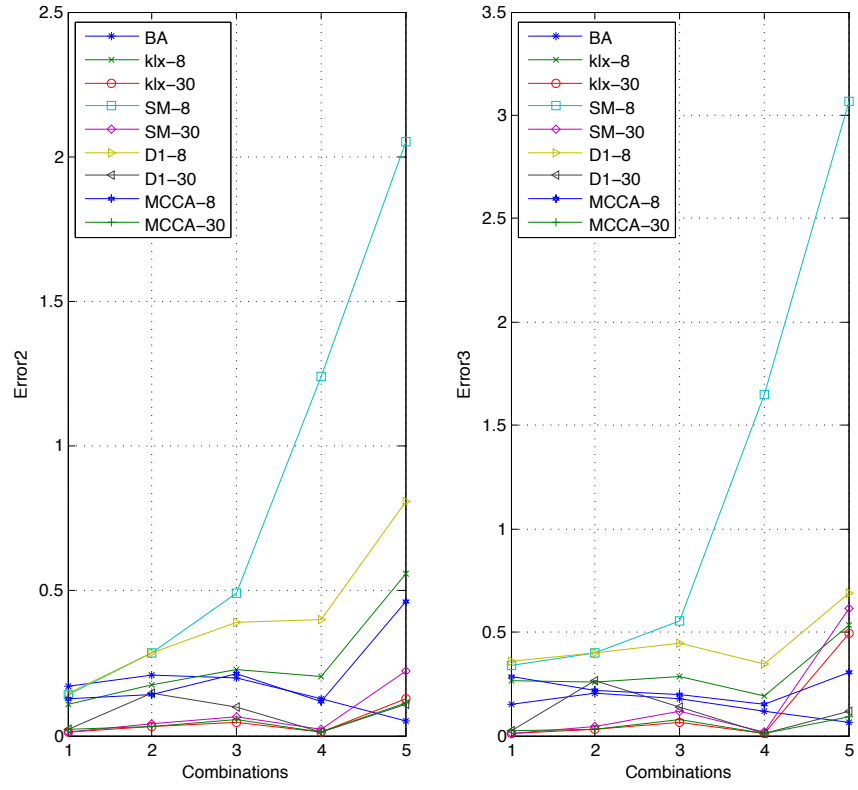


Figure 9.6: $Error2$ and $Error3$ of $\mathcal{E}_{\Theta_X}[i]$, $i=1, \dots, 5$, the i -th DCR-fused BoE.

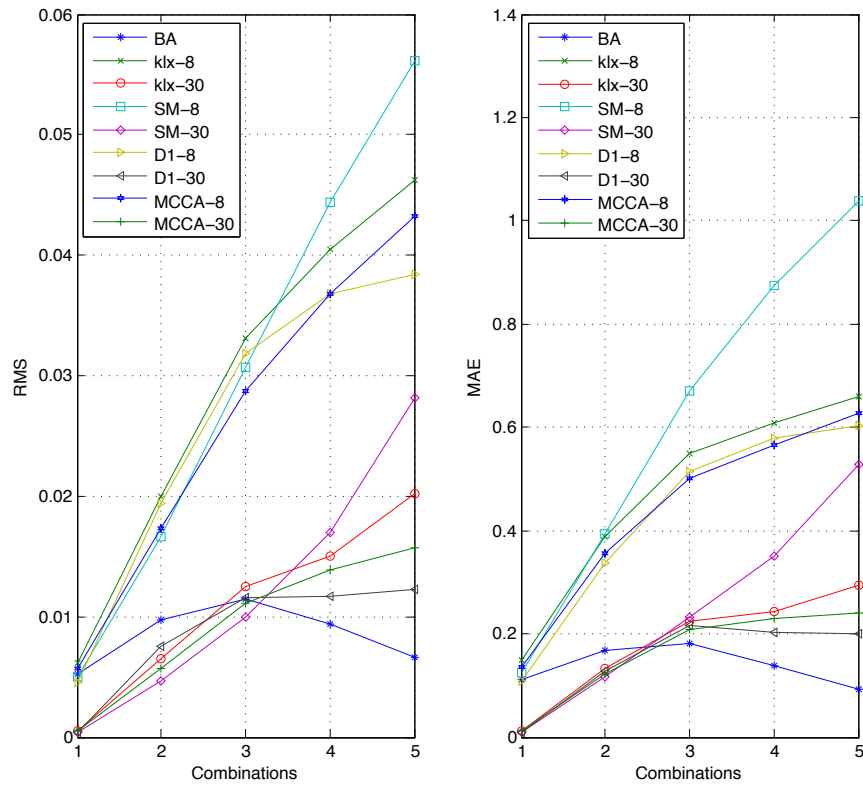


Figure 9.7: RMS and MAE of $\mathcal{E}_{\Theta_X}[i]$, $i = 1, \dots, 5$, the i -th DCR-fused BoE.

Part IV

EPILOGUE

Future Research Directions

THE work presented in this dissertation provides a rather comprehensive theoretical analysis of several core fusion problems. The analytical tools that we have provided—the CFE and its variants, the CCT and associated computational methods, the consensus protocol and their convergence analysis, and the approximations techniques—set a solid foundation for further development of tools that cater to other real-life soft/hard fusion application scenarios. In this chapter, we provide some thought and guidelines that maybe used in order to extend this work or to generate new research based on the provided analytical tools.

10.1 Distributed Fusion

The idea behind this approach is to develop a computationally efficient “joint” representation of the DST data model involving multiple sources. If the sources can be localized based on some “independence” notions, then the computations can be localized and the overall computational costs can be significantly reduced. Let us proceed as follows.

10.1.1 Problem Formulation

Let $\underline{X} = [X_1, X_2, \dots, X_N]$ be a set of sources where each source X_i is modeled via FoD $\Theta_i = \{\theta_{i,k_i}\}$, $k_i=1, |\Theta_i|$ and $i=1, \dots, N$. Let the BoE $\mathcal{E}_\Theta = \{\Theta, \mathcal{F}_\Theta, m_\Theta\}$ denote the joint BoE where $\Theta = \prod_{i=1}^N \Theta_i$. Then we model the BoE \mathcal{E}_Θ as a DST Markov random field (MRF):

Definition 40 (DS-MRF). *A DST Markov random field (DS-MRF) or undirected DST graphical model is a graphical model where Markovian properties (conditional independencies) of random variables are described by an undirected graph. The joint mass function of random variables $\underline{X} = [X_1, X_2, \dots, X_N]$ is defined as*

$$m(\underline{X}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_c(\underline{X}_c),$$

where $Z = \sum_{\underline{x} \in \Theta} \prod_{c \in \mathcal{C}} \psi_c(\underline{x}_c)$ is the partition function and $\psi_c(\underline{X}_c)$ is the potential defined over the clique $c \in \mathcal{C}$. The cliques define the factorization of random variables $[\underline{X}, \underline{Y}]$ according to the conditional dependencies as described by the graph. ■

10.1.2 Distributed Fusion via DS-MRF

Computations in a MRF can be performed in a clique-wise fashion. For instance, take $N = 5$ with $|\Theta_i| = 2$, for $i=1, \dots, 5$. Thus, for instance, a belief computation in the joint space Θ requires $2^{10} = 1024$ computations if there is no factorization. However, if X_1, X_2, X_3 and X_4, X_5 form two separate cliques, then the total number of computations reduces to $2^6 + 2^4 = 80$, providing over 92% computational savings (in terms of the number of computations).

10.1.3 CCT-based message passing

The computational advantages of probabilistic graphical models (e.g., Bayesian networks, Conditional Random Fields, etc.) are due to the efficient message-passing

schemes. The “holy grail” of message passing schemes is precisely the fact that the conditional probability $P(B|A)$ is proportional to $P(A \cap B)$ (from Bayes Theorem). This allows one to compute the joint probabilities and then estimate the conditionals via normalization. However, this approach fails in the case of DST conditionals, such as the FH conditional. Hence, conditionals have to be explicitly computed in a DST message passing scheme. Since, the CCT provides a complete characterization of the conditional core, one could exploit this to implement efficient message passing schemes. Further, the graphical approach to conditional mass computation suggests a recursive implementation of conditionals for composite propositions. For example, propositions with cardinality one (i.e., $|B| = 1$) are computed first, then the propositions with cardinality two, etc. One could also look into incorporating this into the message passing, so that both the conditional masses and inferences can be computed at once.

10.2 Implication Rules as Conditionals

Popularity of DS belief theory in soft/hard applications is mainly due to its flexibility in modeling complex data models with ease. Consider a $2-dof$ (degrees of freedom) implication rule $A \implies B : \Theta_A \mapsto \Theta_B$ with $[\alpha, \beta]$. This cannot be modeled in probability theory, without making simplifying assumptions. On the other hand, DS theory does provide strategies to model these implications; however, the existing models are rather complex and computationally more expensive. In probability theory, the inappropriateness of using conditional probabilities for modeling implications is a well studied and discussed topic (see [54] for a detailed discussion on this). However, in DS theory we can precisely capture the meaning of the $2-dof$ rule by modeling it as,

$$\text{Bl}(B|A) = \alpha; \text{Pl}(B|A) = \beta$$

These models can be easily used with the proposed fusion strategy (CFE) with least computational burden. However, the challenge lies in deriving $Bl(\cdot)$ and $Pl(\cdot)$ on the joint space (i.e., $\Theta_A \times \Theta_B$), where these conditional functions do not possess one unique mapping. The CCT probably provides the best avenue to pursue towards deriving models for implication rules based on conditional notions.

10.3 Reliability/Credibility Estimation

10.3.1 Source Reliability Estimation via ACF

Soft sources (e.g., a human witness) are neither capable of providing evidence in all occasions nor capable of understanding all contexts. Thus, the information provided by these sources are often sparse and highly context dependent. The idea behind using ACF is to make use of contextual information and past experience with soft sources to estimate the reliability of sources for a situation that is currently under observation. Let us explain this approach via an example.

Incident \rightarrow Witness \downarrow	Incident #1	Incident #2	Incident #3	\dots	Incident #M
Witness ₁	0.7	-	0.2	\dots	0.5
Witness ₂	-	0.9	-	\dots	-
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
Witness _N	-	0.9	-	\dots	0.1

Table 10.1: Numerical assessments of truthfulness of evidence

10.3.1.1 Scenario and Problem Formulation

Assume a battlefield scenario, where human witnesses provide evidence regarding suspicious activities in a neighborhood. Once the activity is fully understood, witness statements are assessed by analysts (e.g., a warfare experts) in order to numerically

quantify the truthfulness of the evidence (See Table 10.2.) This determination depends on analyst’s domain expertise, past experience with the witness, context, and other factors. Further, a context analysis of the incident is also performed (See Table 10.2.)

Table 10.2: Context classification of incidents

Context \rightarrow Incident \downarrow	Explosives	Region	Clan X	\dots	Context _M
Incident #1	1	1	0	\dots	0
Incident #2	0	1	0	\dots	1
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
Incident #N	1	1	0	\dots	0

10.3.2 Source Reliability Assessment

When the reliability of evidence sources need to be assessed for an ongoing scenario, (i) the context variables are computed to categorize the ongoing situation, (ii) using historical data in collaboration with other witnesses, ACF is used to estimate the measures of truthfulness of new evidence provided by witnesses making, and finally (iii) the assessments are combined (using an appropriate methodology [104]) in a context-aware setting to estimate the reliability of the source in the context of the ongoing situation.

Conclusion

HOW one may most effectively incorporate soft data into the fusion process has attracted considerable attention in the evidence fusion community. In soft/hard data fusion, the very nature of the soft data introduces several challenges that have been identified as C_1 - C_5 in Chapter 1. We have proposed an analytical framework addressing these challenges associated with soft/hard fusion.

CFE, the evidence fusion strategy proposed in Chapter 5, possesses several intuitively appealing features which seem to indicate its suitability for soft/hard data fusion. CFE addresses the challenges $C_1 - C_4$ via the use of the DST conditional approach. Among the various attractive properties that the use of the conditional approach contributes, it also possesses the ability to handle non-identical FoDs. Our use of the ported conditional notations allows one to easily understand and also to represent fusion operations in both identical and non-identical FoDs via a unified operator.

Consensus is an important fusion problem, where an agreement among a group of agents is sought via repeated exchange of information. This is particularly useful in soft/hard fusion environments where the opinions (soft data received from agents) can often be contradictory and inconsistent. The DST consensus protocol is applicable

to any fusion domain where the data are modeled as DST BoEs. This protocol generates a *rational consensus*, which makes intuitive sense and immune to errors, in the sense that it is capable of “driving” a consensus to an estimate of the GT (Ground Truth), when such evidence is available. The consensus strategy we have developed is applicable to a broad class of networks; and the convergence criteria therein can be easily applied to check for convergence of arbitrary networks.

Data validation is imperative in fusion operations in order to mitigate the risks of making decisions based on questionable fused outputs. However, reliability or other information needed to assess the credibility of data provided by soft sources is often unavailable in many application domains. Based on our notion of rational consensus, we have proposed a credibility estimation method that can be used to self-validate evidence in the absence of such data. These credibility estimates can be used in fusion equations, and can be used for reliability estimation of sources for subsequent fusion operations.

The core of the analytical framework presented here is based on the FH conditional notions, perhaps the most appropriate DST conditional notion for soft/hard fusion operations due to its close connection to Bayesian probability theory. The CCT (Conditional Core Theorem) provides a complete characterization of the FH conditional notions. We believe this theorem is fundamental in understanding various fusion strategies based upon FH conditional notions and the FH conditionals itself. We have also addressed the converse problem via the CCG (conditional core generator) bounds. These bounds provide a way to bound the sets of conditioning events that may have caused a belief change, again without any numerical computations. This result is of significant practical value to an analyst, for instance, in studying the sensitivity of a knowledge base with respect to updates. In addition to providing insight into the conditioning operations, one direct contribution of this important result is that it establishes the theoretical grounds for developing algorithms for efficient computation of DST conditionals.

The strength of DST methods are often hampered by high computational requirements. Computational complexity of even the basic operations, such as conditioning, can grow exponentially with respect to the size of the FoD. Computational complexity of the CCT-based approach for conditional computation does not change as the size of the FoD is increased. Thus, the CCT-based method scales quite well for large FoDs unlike the standard method of conditional computation which can become prohibitive in terms of computational overhead. This is a very significant result, especially in the case of soft data where one often has to deal with large FoDs possessing only a smaller number of focal elements. For swift assessment and analysis of information, the soft/hard fusion framework must be computationally very efficient. The MCMC sampling based method allows for further improvements by approximating the DST BoEs by computationally efficient variants. The advantage of the proposed method is that it allows one to select an appropriate objective function for approximation, and this function can be chosen to suit the application and decision-making criteria.

The work presented in this dissertation provides a rather comprehensive theoretical analysis of several core fusion problems. The analytical tools that we have provided (e.g., the CUE and CFE, the CCT, consensus protocols, their convergence analysis, the approximation techniques, etc.) set a solid foundation for further development of tools that cater to other real-life soft/hard fusion application scenarios.

Appendices

Appendix A

Proofs

A.1 Chapter 4 Proofs

A notion that captures the “size” of a set is

Definition 41 (Largest Set). *Let Ω be any arbitrary set. For an arbitrary $\omega \in \Omega$ and some property $\mathcal{P} : \Omega \mapsto \{\text{true}, \text{false}\}$, we say that ω^* is the largest element in Ω satisfying the property \mathcal{P} and denote it as $\omega^* = \text{Lgst}(\omega \in \Omega \mid \mathcal{P})$, if $\mathcal{P}(\omega^*) = \text{true}$ and $\mathcal{P}(\omega') = \text{false}, \forall \omega' \in \Omega$ s.t. $\omega^* \subset \omega'$. By convention, we take $\omega^* = \emptyset$, if $\mathcal{P}(\omega) = \text{false}, \forall \omega \in \Omega$.* □

A.1.1 A Useful Set Construction

The proof of the CCT requires consideration of several cases. We will make use of a certain set construction that captures all these cases in one setting, thus enabling us to provide a general proof. An important result regarding this particular collection of sets is Corollary 27. To establish Corollary 27, we make use of the following identity the dual of which can be thought of forming the basis for the popular *inclusion-exclusion principle*.

Claim 26. Consider an arbitrary collection of sets A_1, \dots, A_n , s.t. $A = \bigcup_{i=1}^n A_i$. Then, the identity

$$1_{\cap A} = \sum_{k=1}^n (-1)^{k-1} \sum_{\substack{\mathcal{I} \subseteq \{1, \dots, n\}; \\ |\mathcal{I}|=k}} 1_{A_{\cup \mathcal{I}}}$$

holds true for identity functions $1_{(\cdot)}$, where $A_{\cup \mathcal{I}} = \bigcup_{i \in \mathcal{I}} A_i$. □

Proof. Easily established via mathematical induction. ■

Corollary 27. For a given set $B \subseteq \Theta$, consider any arbitrary collection of sets $\{B_k\}$ s.t. $B_k \subset B$, $k = \overline{1, N}$, and $\bigcup_{j=1}^k B \setminus B_j \subset B_{k+1} \subset B$, $k = \overline{1, N-1}$. Then

$$\begin{aligned} \text{Bl}_{\Theta}(B) &= (3 - 2^N) \cdot \text{Bl}_{\Theta}(\mathbf{B}_{\cap}^{(N)}) + \sum_{(N, \mathcal{J})} (-1)^{|\mathcal{J}|-1} \cdot \text{Bl}_{\Theta}(\mathbf{B}_{\cap}^{(N)} \cup \mathbf{B}_B^{(\mathcal{J})}) \\ &\quad + \sum_{\substack{\emptyset \neq Y_k \subseteq B \setminus B_k; \\ k \in \{1, \dots, N\}}} \left\{ m_{\Theta}(\mathbf{Y}_{\cup}^{(N)}) + \sum_{\emptyset \neq X \subseteq \mathbf{B}_{\cap}^{(N)}} m_{\Theta}(X \cup \mathbf{Y}_{\cup}^{(N)}) \right\}, \end{aligned} \tag{A.1}$$

where we use the notation

$$\begin{aligned} \mathbf{B}_{\cap}^{(N)} &= \bigcap_{i=1}^N B_i; \quad \mathbf{Y}_{\cup}^{(N)} = \bigcup_{j=1}^N Y_j; \quad \mathbf{B}_B^{(\mathcal{J})} = \bigcup_{j \in \mathcal{J}} B \setminus B_j; \\ \sum_{(N, \mathcal{J})} &= \sum_{n=1}^{N-1} \sum_{\ell=1}^n \sum_{\substack{\mathcal{I} \subseteq \{1, \dots, N\}; \\ |\mathcal{I}|=n}} \sum_{\substack{\mathcal{J} \subseteq \mathcal{I}; \\ |\mathcal{J}|=\ell}}, \end{aligned}$$

and $\text{Bl}_{\Theta}(\cdot) : 2^{\Theta} \mapsto [0, 1]$ is a valid belief function on Θ . □

Proof. By construction, we have

$$B = \mathbf{B}_{\cap}^{(N)} \cup \mathbf{B}_B^{(\overline{1, N})}; \quad \mathbf{B}_{\cap}^{(N)} \cap \mathbf{B}_B^{(\overline{1, N})} = \emptyset.$$

Thus, we can expand $\text{Bl}_{\Theta}(B)$ as

$$\text{Bl}_{\Theta}(B) = \text{Bl}_{\Theta} \left[\mathbf{B}_{\cap}^{(N)} \cup \mathbf{B}_B^{(\overline{1, N})} \right]$$

$$= \text{Bl}_\Theta(\mathbf{B}_\cap^{(N)}) + \text{Bl}_\Theta(\mathbf{B}_B^{(\overline{1,N})}) + \sum_{\substack{\emptyset \neq X \subseteq \mathbf{B}_\cap^{(N)}; \\ \emptyset \neq Y \subseteq \mathbf{B}_B^{(\overline{1,N})}}} m_\Theta(X \cup Y). \quad (\text{A.2})$$

Since $(B \setminus B_i) \cap (B \setminus B_j) = \emptyset$, $\forall i \neq j$, we can expand the second belief term as

$$\begin{aligned} \mathbf{B}_B^{(\overline{1,N})} &= \sum_{\emptyset \neq Y \subseteq \mathbf{B}_B^{(\overline{1,N})}} m_\Theta(Y) + m_\Theta(\emptyset) \\ &= \sum_{n=1}^{N-1} \sum_{\substack{\mathcal{I} \subset \{1, \dots, N\}; \\ |\mathcal{I}|=n}} \sum_{\substack{\emptyset \neq Y_k \subseteq B \setminus B_k; \\ k \in \mathcal{I}}} m_\Theta\left(\bigcup_{i \in \mathcal{I}} Y_i\right) + \sum_{\substack{\emptyset \neq Y_k \subseteq B \setminus B_k; \\ k \in \{1, \dots, N\}}} m_\Theta(\mathbf{Y}_\cup^{(N)}). \end{aligned}$$

Now, use Claim 26 on $\bigcup_{i \in \mathcal{I}} Y_i$, $Y_k \subseteq B \setminus B_k$, $k \in \mathcal{I}$, $\forall \mathcal{I} \subset \{1, \dots, N\}$, to get

$$\begin{aligned} \mathbf{B}_B^{(\overline{1,N})} &= \sum_{n=1}^{N-1} \sum_{\substack{\mathcal{I} \subset \{1, \dots, N\}; \\ |\mathcal{I}|=n}} \sum_{l=1}^n (-1)^{l-1} \sum_{\substack{\mathcal{J} \subseteq \mathcal{I}; \\ |\mathcal{J}|=l}} \sum_{\substack{\emptyset \neq Y \subseteq \bigcup_{j \in \mathcal{J}} B \setminus B_j}} m_\Theta(Y) + \sum_{\substack{\emptyset \neq Y_k \subseteq B \setminus B_k; \\ k \in \{1, \dots, N\}}} m_\Theta(\mathbf{Y}_\cup^{(N)}) \\ &= \sum_{(N, \mathcal{J})} (-1)^{|\mathcal{J}|-1} \sum_{\emptyset \neq Y \subseteq \mathbf{B}_B^{(\mathcal{J})}} m_\Theta(Y) + \sum_{\substack{\emptyset \neq Y_k \subseteq B \setminus B_k; \\ k \in \{1, \dots, N\}}} m_\Theta(\mathbf{Y}_\cup^{(N)}). \quad (\text{A.3}) \end{aligned}$$

Similarly, expand the second summation in (A.2) and substitute back into (A.2)

with (A.3) to get

$$\begin{aligned} &\text{Bl}_\Theta(B) \\ &= \text{Bl}_\Theta(\mathbf{B}_\cap^{(N)}) + \sum_{(N, \mathcal{J})} (-1)^{|\mathcal{J}|-1} \cdot \left[\sum_{\emptyset \neq Y \subseteq \mathbf{B}_B^{(\mathcal{J})}} m_\Theta(Y) + \sum_{\substack{\emptyset \neq X \subseteq \mathbf{B}_\cap^{(N)}; \\ \emptyset \neq Y \subseteq \mathbf{B}_B^{(\mathcal{J})}}} m_\Theta(X \cup Y) \right] \\ &\quad + \sum_{\substack{\emptyset \neq Y_k \subseteq B \setminus B_k; \\ k \in \{1, \dots, N\}}} m_\Theta(\mathbf{Y}_\cup^{(N)}) + \sum_{\substack{\emptyset \neq X \subseteq \mathbf{B}_\cap^{(N)}; \\ \emptyset \neq Y_k \subseteq B \setminus B_k; \\ k \in \{1, \dots, N\}}} m_\Theta(X \cup \mathbf{Y}_\cup^{(N)}) \\ &= \text{Bl}_\Theta(\mathbf{B}_\cap^{(N)}) + \sum_{(N, \mathcal{J})} (-1)^{|\mathcal{J}|-1} \left[\text{Bl}_\Theta(\mathbf{B}_\cap^{(N)} \cup \mathbf{B}_B^{(\mathcal{J})}) - \text{Bl}_\Theta(\mathbf{B}_\cap^{(N)}) \right] \\ &\quad + \sum_{\substack{\emptyset \neq Y_k \subseteq B \setminus B_k; \\ k \in \{1, \dots, N\}}} \left[m_\Theta(\mathbf{Y}_\cup^{(N)}) + \sum_{\emptyset \neq X \subseteq \mathbf{B}_\cap^{(N)}} m_\Theta(X \cup \mathbf{Y}_\cup^{(N)}) \right] \end{aligned}$$

$$\begin{aligned}
&= \left[\sum_{(N, \mathcal{J})} (-1)^{|\mathcal{J}|} + 1 \right] \text{Bl}_{\Theta}(\mathbf{B}_{\cap}^{(N)}) + \sum_{(N, \mathcal{J})} (-1)^{|\mathcal{J}|-1} \left[\text{Bl}_{\Theta}(\mathbf{B}_{\cap}^{(N)} \cup \mathbf{B}_B^{(\mathcal{J})}) \right] \\
&\quad + \sum_{\substack{\emptyset \neq Y_k \subseteq B \setminus B_k; \\ k \in \{1, \dots, N\}}} \left[m_{\Theta}(\mathbf{Y}_{\cup}^{(N)}) + \sum_{\emptyset \neq X \subseteq \mathbf{B}_{\cap}^{(N)}} m_{\Theta}(X \cup \mathbf{Y}_{\cup}^{(N)}) \right].
\end{aligned}$$

The binomial theorem yields

$$\begin{aligned}
\sum_{(N, \mathcal{J})} (-1)^{|\mathcal{J}|} + 1 &= \sum_{n=1}^{N-1} \binom{N}{n} 1^n \cdot \sum_{\ell=1}^n \binom{n}{\ell} (-1)^{\ell} + 1 \\
&= [2^N - 2] \cdot (-1) + 1 = 3 - 2^N.
\end{aligned}$$

Substitute this in the $\text{Bl}_{\Theta}(B)$ expansion above to obtain the required result. ■

A.1.2 Proof of the CCT

Proof. Given $\mathcal{E} = \{\Theta, \mathfrak{F}_{\Theta}, m_{\Theta}(\cdot)\}$ and $A \in \hat{\mathfrak{F}}_{\Theta}$, we have to prove the following:

$$\begin{aligned}
m_{\Theta}(B|A) > 0 &\iff \exists X \in \text{in}(A), \exists Y \in \text{OUT}(A) \cup \{\emptyset\}, \\
&\text{s.t. } B = X \cup Y.
\end{aligned}$$

We need to consider three cases.

CASE 1. $B \not\subseteq A$:

In this case, B cannot be expressed as $B = X \cup Y$, $X \in \text{in}(A)$, $Y \in \text{OUT}(A) \cup \{\emptyset\}$.

Also, $\text{Bl}_{\Theta}(B|A) = \text{Bl}_{\Theta}(B \cap A|A)$ implies $m_{\Theta}(B|A) = 0$, $\forall B \not\subseteq A$. Hence, the CCT holds true in this case. ■

CASE 2. $B \subseteq A$ and $B \in \mathfrak{F}_{\Theta}$:

Such a B can indeed be expressed as $B = X \cup Y$ by taking $X = B$, $Y = \emptyset$. Express

the expression in Claim ?? in masses to get

$$\sum_{C \subseteq B} m_{\Theta}(C|A) = \sum_{C \subseteq B} \frac{m_{\Theta}(C)}{\text{Pl}_{\Theta}(A) - \mathcal{S}(\bar{A}; B)} \geq \sum_{C \subseteq B} \frac{m_{\Theta}(C)}{\text{Pl}_{\Theta}(A) - \mathcal{S}(\bar{A}; C)},$$

because $\mathcal{S}(\bar{A}; C) \leq \mathcal{S}(\bar{A}; B)$ whenever $C \subseteq B$. Since this holds true for all $B \subseteq \Theta$, we have

$$m_{\Theta}(B|A) \geq \frac{m_{\Theta}(B)}{\text{Pl}_{\Theta}(A) - \mathcal{S}(\bar{A}; B)} > 0,$$

because $m_{\Theta}(B) > 0$, $B \in \mathfrak{F}_{\Theta}$. Hence, the CCT holds true in this case. \blacksquare

CASE 3. $B \subseteq A$ and $B \notin \mathfrak{F}_{\Theta}$:

With $B \notin \mathfrak{F}_{\Theta}$, we must have $Y \neq \emptyset$. Thus, we only have to prove the following:

$$m_{\Theta}(B|A) > 0 \iff \exists X \in \text{in}(A), \exists Y \in \text{OUT}(A), \text{ s.t. } B = X \cup Y.$$

With no loss of generality, from now on, we can restrict our attention to $X \in \text{in}(A)$, $X \subseteq B$, and $Y \in \text{OUT}(A)$, $Y \subseteq B$; it is impossible to represent B as $X \cup Y$ otherwise. Let us consider the ‘forward’ and ‘reverse’ directions separately.

Case 3.a. ‘Forward’ Direction:

Here, given $B \notin \mathfrak{F}_{\Theta}$ and $B \subseteq A$, we need to prove

$$m_{\Theta}(B|A) > 0 \implies \exists X \in \text{in}(A), X \subseteq B, \exists Y \in \text{OUT}(A), Y \subseteq B, \text{ s.t. } B = X \cup Y.$$

Let us prove the contrapositive. Note that under the condition $X \subseteq B$ and $Y \subseteq B$, if $B \neq X \cup Y$, we must have $(X \cup Y) \subset B$, with $X \subset B$ and $Y \subset B$. Thus, the contrapositive is

$$(X \cup Y) \subset B, \forall X \in \text{in}(A) \text{ s.t. } X \subset B \text{ and } \forall Y \in \text{OUT}(A) \text{ s.t. } Y \subset B \implies m_{\Theta}(B|A) = 0. \quad (\text{A.4})$$

We proceed as follows. Pick $\hat{B} = \hat{X} \cup \hat{Y}$, where

$$\hat{X} = \bigcup_{\substack{X \in \text{in}(A); \\ X \subset B}} X; \quad \hat{Y} = \text{Lgst} \{Y \in \text{OUT}(A) \mid Y \subset B\}.$$

Notice that $\mathcal{S}(\bar{A}; B) = \mathcal{S}(\bar{A}; \hat{Y})$. Further, $\hat{B} \subseteq B$, since $\hat{X} \subseteq B$ and $\hat{Y} \subset B$. So, we have two possibilities to consider: $\hat{B} \subset B$ and $\hat{B} = B$.

Case 3.a.i. ‘Forward’ Direction, Part 1. $\hat{B} = (\hat{X} \cup \hat{Y}) \subset B$: Here we have $\mathcal{S}(\bar{A}; \hat{B}) = \mathcal{S}(\bar{A}; \hat{Y}) = \mathcal{S}(\bar{A}; B)$ and $\text{Bl}_\Theta(B) = \text{Bl}_\Theta(\hat{X}) = \text{Bl}_\Theta(\hat{B})$ because $B \notin \mathfrak{F}_\Theta$ and \hat{X} capture all the focal elements in B . So,

$$\begin{aligned} \text{Bl}_\Theta(B|A) &= \frac{\text{Bl}_\Theta(B)}{\text{Pl}_\Theta(A) - \mathcal{S}(\bar{A}; B)} \\ &= \frac{\text{Bl}_\Theta(\hat{B})}{\text{Pl}_\Theta(A) - \mathcal{S}(\bar{A}; \hat{B})} = \text{Bl}_\Theta(\hat{B}|A), \end{aligned}$$

Thus, we must have $m_\Theta(G|A) = 0, \forall \hat{B} \subset G \subseteq B$. In particular, $m_\Theta(B|A) = 0$.

Hence, the contrapositive given in (A.4) holds true for $\hat{B} \subset B$. ■

Case 3.a.ii. ‘Forward’ Direction, Part 2. $\hat{B} = \hat{X} \cup \hat{Y} = B$: This case requires a little more effort to establish. We proceed with the following set construction.

Let $B_1 = X_1 \cup \hat{Y}$, where

$$X_1 = \text{Lgst} \left\{ X \in \text{IN}(A) \mid X \cup \hat{Y} \subset B \right\}.$$

We must have $X_1 \neq \emptyset$ because, given that $\exists X \in \text{in}(A), X \subset B$, s.t. $X \cup \hat{Y} \subset B$ (and trivially, $\text{in}(A) \subseteq \text{IN}(A)$). So, $\mathcal{S}(\bar{A}; B) = \mathcal{S}(\bar{A}; B_1)$ since $\hat{Y} \subseteq B_1 \subset B$. Note that,

X_1 exhausts all the focal elements in B s.t. $X_1 \cup \hat{Y} \subset B$. Thus, B cannot contain focal elements that are (i) strict subsets of $B \setminus B_1$; or (ii) unions of a subset of B_1 and a strict subset of $B \setminus B_1$. So, we must have $m_\Theta(C) = 0$, for all $C = X \cup Y$ s.t. $X' \subset B \setminus B_1$ and $Y \subseteq B_1$.

We may now continue to form $B_k = X_k \cup \hat{Y}$, with $X_k \neq \emptyset$ for $k \geq 2$, where

$$X_k = \text{Lgst} \left\{ X \in \text{IN}(A) \mid \mathbf{B}_B^{(\overline{1, k-1})} \subseteq X \text{ s.t. } X \cup \hat{Y} \subset B \right\},$$

and terminate at the smallest N s.t. $X_{N+1} = \emptyset$. Clearly, for this N , we must have

$$m_\Theta(C) = 0, \quad \forall C \text{ s.t. } \mathbf{B}_B^{(\overline{1, N})} \subseteq C \subset B. \quad (\text{A.5})$$

Note that, $B_1 = X_1 \cup \hat{Y} \subset B = \hat{X} \cup \hat{Y}$. Hence, $\exists X^* \in \text{in}(A)$, $X^* \subset B$, s.t. $(X_1 \cup X^*) \cup \hat{Y} = B$; conversely, $\exists X^* \in \text{in}(A)$ s.t. $B \setminus B_1 \subseteq X^*$. So, we must have $X_2 \neq \emptyset$. Hence, $N \geq 2$. Moreover, this procedure is guaranteed to terminate with an upper bound of $N \leq |B \setminus \hat{Y}|$ (see Appendix A.1.3 for the proof).

By construction, the sets B_k satisfy the following properties:

$$\hat{Y} \subseteq B_k \subset B, \quad k = \overline{1, N}; \quad (\text{A.6})$$

$$\mathcal{S} \left(\overline{A}; \bigcap_{i \in \mathcal{I}} B_i \right) = \mathcal{S}(\overline{A}; B), \quad I \subseteq \{1, \dots, N\}; \quad (\text{A.7})$$

$$\mathbf{B}_B^{(\overline{1, k})} \subset B_{k+1} \subset B, \quad k = \overline{1, N-1}; \quad (\text{A.8})$$

$$m_\Theta(C) = 0, \quad \forall C = X \cup Y \text{ s.t.}$$

$$X \subseteq \mathbf{B}_\cap^{(k)}, \quad Y \neq \mathbf{B}_B^{(\overline{1, k-1})} \text{ and}$$

$$\emptyset \neq Y \subset \mathbf{B}_B^{(\mathcal{I})}, \quad \mathcal{I} \subset \{1, \dots, k\} \quad k = \overline{1, N}. \quad (\text{A.9})$$

With this set construction in place, we may now proceed as follows to establish the required result.

First notice that, Corollary 27 is applicable to the collection of sets $\{B_k\}_{k=1}^N$. For the expression for $\text{Bl}_\Theta(B)$, consider the term corresponding to the last term in (A.1):

$$\begin{aligned}
& \sum_{\substack{\emptyset \neq Y_k \subseteq B \setminus B_k; \\ k \in \{1, \dots, N\}}} \left[m_\Theta(\mathbf{Y}_\cup^{(N)}) + \sum_{\emptyset \neq X \subseteq \mathbf{B}_\cap^{(N)}} m_\Theta(X \cup \mathbf{Y}_\cup^{(N)}) \right] \\
&= m_\Theta(\mathbf{B}_B^{(\overline{1}, N)}) + \sum_{\emptyset \neq X \subseteq \mathbf{B}_\cap^{(N)}} m_\Theta(X \cup \mathbf{B}_B^{(\overline{1}, N)}) \\
&\quad + \sum_{\substack{\emptyset \neq Y_k \subseteq B \setminus B_k; \\ k \in \{1, \dots, N\}}} \left[m_\Theta(\mathbf{Y}_\cup^{(N)}) + \sum_{\emptyset \neq X \subseteq \mathbf{B}_\cap^{(N)}} m_\Theta(X \cup \mathbf{Y}_\cup^{(N)}) \right] \\
&= \sum_{\mathbf{B}_B^{(\overline{1}, N)} \subseteq C \subseteq B} m_\Theta(C) + m_\Theta(B).
\end{aligned}$$

Note that all the terms in the last summation group are zero from property (A.9).

Using property (A.5) from the termination criterion of set generation and the fact that $B \notin \mathfrak{F}$, we get

$$\sum_{\substack{\emptyset \neq Y_k \subseteq B \setminus B_k; \\ k \in \{1, \dots, N\}}} \left[m_\Theta(\mathbf{Y}_\cup^{(N)}) + \sum_{\emptyset \neq X \subseteq \mathbf{B}_\cap^{(N)}} m_\Theta(X \cup \mathbf{Y}_\cup^{(N)}) \right] = 0.$$

So, $\text{Bl}_\Theta(B)$ reduces to

$$\text{Bl}_\Theta(B) = (3 - 2^N) \cdot \text{Bl}_\Theta(\mathbf{B}_\cap^{(N)}) + \sum^{(N, \mathcal{J})} (-1)^{|\mathcal{J}|-1} \cdot \text{Bl}_\Theta(\mathbf{B}_\cap^{(N)} \cup \mathbf{B}_B^{(\mathcal{J})}). \quad (\text{A.10})$$

Use this and property (A.7) in Claim ?? to get

$$\text{Bl}_\Theta(B|A) = (3 - 2^N) \cdot \text{Bl}_\Theta(\mathbf{B}_\cap^{(N)}|A) + \sum^{(N, \mathcal{J})} (-1)^{|\mathcal{J}|-1} \cdot \text{Bl}_\Theta(\mathbf{B}_\cap^{(N)} \cup \mathbf{B}_B^{(\mathcal{J})}|A). \quad (\text{A.11})$$

Next, apply Corollary 27 to the collection $\{B_k\}$ to obtain

$$\text{Bl}_\Theta(B|A) = (3 - 2^N) \cdot \text{Bl}_\Theta(\mathbf{B}_\cap^{(N)}|A) + \sum^{(N, \mathcal{J})} (-1)^{|\mathcal{J}|-1} \cdot \text{Bl}_\Theta(\mathbf{B}_\cap^{(N)} \cup \mathbf{B}_B^{(\mathcal{J})}|A)$$

$$+ \sum_{\substack{\emptyset \neq Y_k \subseteq B \setminus B_k; \\ k \in \{1, \dots, N\}}} \left[m_{\Theta}(\mathbf{Y}_{\cup}^{(N)} | A) + \sum_{\emptyset \neq X \subseteq \mathbf{B}_{\cap}^{(N)}} m_{\Theta}(X \cup \mathbf{Y}_{\cup}^{(N)} | A) \right]. \quad (\text{A.12})$$

Now, equate terms in (A.11) and (A.12) to get

$$\sum_{\substack{\emptyset \neq Y_k \subseteq B \setminus B_k; \\ k \in \{1, \dots, N\}}} \left[m_{\Theta}(\mathbf{Y}_{\cup}^{(N)} | A) + \sum_{\emptyset \neq X \subseteq \mathbf{B}_{\cap}^{(N)}} m_{\Theta}(X \cup \mathbf{Y}_{\cup}^{(N)} | A) \right] = 0.$$

Since $m(\cdot | A) \geq 0$, each individual mass term must equal to zero. In particular,

$$m_{\Theta}(\mathbf{B}_{\cap}^{(N)} \cup \mathbf{B}_B^{(\overline{1}, N)} | A) = m_{\Theta}(B | A) = 0.$$

Thus, the contrapositive in (A.4) holds true for $\hat{B} = B$. ■

Thus the ‘Forward’ direction of Case 3 is established. ■

Case 3.b. ‘Reverse’ Direction:

Given $B \notin \mathfrak{F}$ and $B \subseteq A$, we now need to prove

$$\exists X \in \text{in}(A), X \subseteq B \text{ and } \exists Y \in \text{OUT}(A), Y \subseteq B \text{ s.t. } B = X \cup Y \implies m_{\Theta}(B | A) > 0.$$

Since $B \notin \mathfrak{F}$, we must have $B \neq X, \forall X \in \text{in}(A)$. Hence, given $B = X \cup Y$ for some $X \in \text{in}(A)$ s.t. $X \subset B$ and some $Y \in \text{OUT}(A)$ s.t. $Y \subseteq B$, we need to prove that $m_{\Theta}(B | A) > 0$. Let us proceed as follows.

Let $G_1 = X_1 \cup Y_1$, where

$$X_1 = \text{Lgst} \{X' \in \text{IN}(A) \mid X \subseteq X' \subset B\};$$

$$Y_1 = \text{Lgst} \{Y' \in \text{OUT}(A) \mid Y \subseteq Y' \text{ s.t. } X_1 \cup Y' \subset B\}.$$

Clearly, $X_1 \neq \emptyset$ and also $G_1 \subset B$. Further, note that $\mathcal{S}(\overline{A}; G_1) < \mathcal{S}(\overline{A}; B)$. Let

$$\hat{Y} = \text{Lgst} \{Y' \in \text{OUT}(A) \mid Y \subseteq Y' \text{ s.t. } X \cup Y' = B\}.$$

Now, $\mathcal{S}(\overline{A}; B) = \mathcal{S}(\overline{A}; \hat{Y})$, since \hat{Y} exhausts all $Y \in \text{OUT}(A)$ s.t. $Y \subseteq B$. Similarly, $\mathcal{S}(\overline{A}; G_1) = \mathcal{S}(\overline{A}; Y_1)$. But, $X \subseteq X_1$ and $X_1 \cup Y_1 \subset X \cup \hat{Y}$. Hence, we must have $Y_1 \subset \hat{Y}$. Thus, $\mathcal{S}(\overline{A}; Y_1) < \mathcal{S}(\overline{A}; \hat{Y})$. Therefore, $\mathcal{S}(\overline{A}; G_1) < \mathcal{S}(\overline{A}; B)$.

We may now continue to form $G_k = X_k \cup Y_k$, with $X_k \neq \emptyset$ for $k \geq 2$, where

$$\begin{aligned} X_k &= \text{Lgst} \left\{ X' \in \text{IN}(A) \mid \mathbf{G}_B^{(\overline{1, k-1})} \subseteq X' \text{ s.t. } X' \subset B \right\}; \\ Y_k &= \text{Lgst} \left\{ Y' \in \text{OUT}(A) \mid X_k \cup Y' \subset B \right\}, \end{aligned}$$

and terminate at the smallest M s.t. $X_{M+1} = \emptyset$. Clearly, for this M , we must have

$$m_\Theta(C) = 0, \forall C \text{ s.t. } \mathbf{G}_B^{(\overline{1, M})} \subseteq C \subset B. \quad (\text{A.13})$$

Moreover, this procedure is guaranteed to terminate with an upper bound of $M < |B|$ (see Appendix A.1.4 for the proof).

Now, consider the sets of the form $C = D \cup E$ s.t. $\emptyset \neq D \subset \mathbf{G}_B^{(\mathcal{I})}$, where $D \neq \mathbf{G}_B^{(\overline{1, k-1})}$, $\mathcal{I} \subset \{1, \dots, k\}$, $E \subseteq \mathbf{G}_\cap^{(k)}$, and $k = \overline{1, M}$. By construction of G_k , $k = \overline{1, M}$, we have $C \notin \mathfrak{F}$ and thus $m_\Theta(C) = 0$. Further, $C \neq X' \cup Y'$, for all $X' \in \text{IN}(A)$, $X' \subseteq C$, and $Y' \in \text{OUT}(A)$, $Y' \subseteq C$. We must therefore have $C \neq X'' \cup Y$, $\forall X'' \in \text{in}(A)$, because $\text{in}(A) \subseteq \text{IN}(A)$. But, we have already proven that $m_\Theta(C|A) = 0$ for all such C .

In summary, the sets G_k satisfy the following properties (first three by construction and the last property as argued above):

$$G_k \subset B, \quad k = \overline{1, M}; \quad (\text{A.14})$$

$$\mathcal{S}(\bar{A}; \mathbf{G}_{\cap}^{(k)}) < \mathcal{S}(\bar{A}; B), \quad k = \overline{1, M-1}; \quad (\text{A.15})$$

$$\mathbf{G}_B^{(\overline{1, k})} \subset G_{k+1} \subset B, \quad k = \overline{1, M-1}; \quad (\text{A.16})$$

$$m_{\Theta}(C) = m_{\Theta}(C|A) = 0,$$

$$\forall C = X \cup Y \text{ s.t. } X \subseteq \mathbf{G}_{\cap}^{(k)},$$

$$\emptyset \neq Y \subset \mathbf{G}_B^{(\mathcal{I})}, \mathcal{I} \subset \{1, \dots, k\}$$

$$\text{and } Y \neq \mathbf{G}_B^{(\overline{1, k-1})}, \quad k = \overline{1, M}. \quad (\text{A.17})$$

With this set construction in place, we may now proceed as follows to establish the required result.

First notice that Corollary 27 is applicable to the collection of sets $\{G_k\}_{k=1}^M$. Similar to (A.10), let us use Claim ?? to express $\text{Bl}_{\Theta}(B|A)$ as

$$\text{Bl}_{\Theta}(B|A) = (3 - 2^M) \cdot \frac{\text{Bl}_{\Theta}(\mathbf{G}_{\cap}^{(M)})}{\text{Pl}_{\Theta}(A) - \mathcal{S}(\bar{A}; B)} + \sum^{(M, \mathcal{J})} (-1)^{|\mathcal{J}|-1} \cdot \frac{\text{Bl}_{\Theta}(\mathbf{G}_{\cap}^{(M)} \cup \mathbf{G}_B^{(\mathcal{J})})}{\text{Pl}_{\Theta}(A) - \mathcal{S}(\bar{A}; B)}.$$

In arriving at this expression, we have used the property (A.17) and the termination condition (A.13). Now, using the property (A.15), one can clearly see that the following inequality holds:

$$\begin{aligned} & (3 - 2^M) \cdot \frac{\text{Bl}_{\Theta}(\mathbf{G}_{\cap}^{(M)})}{\text{Pl}_{\Theta}(A) - \mathcal{S}(\bar{A}; B)} \\ & + \sum^{(M, \mathcal{J})} (-1)^{|\mathcal{J}|-1} \cdot \frac{\text{Bl}_{\Theta}(\mathbf{G}_{\cap}^{(M)} \cup \mathbf{G}_B^{(\mathcal{J})}|A)}{\text{Pl}_{\Theta}(A) - \mathcal{S}(\bar{A}; B)} \\ & > (3 - 2^M) \cdot \frac{\text{Bl}_{\Theta}(\mathbf{G}_{\cap}^{(M)})}{\text{Pl}_{\Theta}(A) - \mathcal{S}(\bar{A}; \mathbf{G}_{\cap}^{(M)})} \\ & + \sum^{(M, \mathcal{J})} (-1)^{|\mathcal{J}|-1} \cdot \frac{\text{Bl}_{\Theta}(\mathbf{G}_{\cap}^{(M)} \cup \mathbf{G}_B^{(\mathcal{J})})}{\text{Pl}_{\Theta}(A) - \mathcal{S}(\bar{A}; \mathbf{G}_{\cap}^{(M)} \cup \mathbf{G}_B^{(\mathcal{J})})}. \end{aligned}$$

So, we get

$$\text{Bl}_{\Theta}(B|A) > (3 - 2^M) \cdot \text{Bl}_{\Theta}(\mathbf{G}_{\cap}^{(M)}|A) + \sum^{(M, \mathcal{J})} (-1)^{|\mathcal{J}|-1} \cdot \text{Bl}_{\Theta}(\mathbf{G}_{\cap}^{(M)} \cup \mathbf{G}_B^{(\mathcal{J})}|A). \quad (\text{A.18})$$

Similarly, we get

$$\begin{aligned}
& \text{Bl}_\Theta(B|A) \\
&= (3 - 2^M) \cdot \text{Bl}_\Theta(\mathbf{G}_\cap^{(M)}|A) + \sum^{(M, \mathcal{J})} (-1)^{|\mathcal{J}|-1} \cdot \text{Bl}_\Theta(\mathbf{G}_\cap^{(M)} \cup \mathbf{G}_B^{(\mathcal{J})}|A) \\
&\quad + \sum_{\substack{\emptyset \neq Y_k \subseteq B \setminus B_k; \\ k \in \{1, \dots, N\}}} \left[m_\Theta(\mathbf{Y}_\cup^{(N)}|A) + \sum_{\emptyset \neq X \subseteq \mathbf{B}_\cap^{(N)}} m_\Theta(X \cup \mathbf{Y}_\cup^{(N)}|A) \right] \\
&= (3 - 2^M) \cdot \text{Bl}_\Theta(\mathbf{G}_\cap^{(M)}|A) + \sum^{(M, \mathcal{J})} (-1)^{|\mathcal{J}|-1} \cdot \text{Bl}_\Theta(\mathbf{G}_\cap^{(M)} \cup \mathbf{G}_B^{(\mathcal{J})}|A) + m_\Theta(B|A),
\end{aligned} \tag{A.19}$$

where all the other terms in the last two summation groups vanish according to property (A.17).

Now, combine (A.19) and (A.18) to conclude that $m(B|A) > 0$.

Thus the ‘Reverse’ direction of Case 3 is established. ■

This completes the proof of the CCT. ■

A.1.3 Proof of Upper Bound for N

Proof. By construction of B_i , $i = \overline{1, N}$, we have

$$\mathbf{B}_\cap^{(N)} \supseteq \hat{Y}; \quad (B \setminus B_i) \cap (B \setminus B_j) = \emptyset, \quad \forall i \neq j.$$

Now,

$$\left| \mathbf{B}_B^{(\overline{1, N})} \right| = \left| B \setminus \mathbf{B}_\cap^{(N)} \right| \leq \left| B \setminus \hat{Y} \right|.$$

By construction, we also have

$$\left| \mathbf{B}_B^{(N)} \right| = \sum_{i=1}^N |B \setminus B_i| \geq N,$$

because $|B \setminus B_i| \geq 1$, when $B_i \subset B$, $\forall i = \overline{1, N}$. Therefore, we have $N \leq |B \setminus \hat{Y}|$. ■

A.1.4 Proof of Upper Bound for M

Proof. By construction of G_i , $i = \overline{1, M}$, we have

$$(B \setminus G_i) \cap (B \setminus G_j) = \emptyset, \forall i \neq j,$$

and

$$\left| \mathbf{G}_B^{(M)} \right| = \sum_{i=1}^M |B \setminus G_i| \geq M,$$

because $|B \setminus G_i| \geq 1$, when $G_i \subset B$, $\forall i = \overline{1, M}$. Trivially, $\left| \mathbf{G}_B^{(M)} \right| < |B|$, so that we have $M < |B|$. ■

A.2 Chapter 5 Proofs

A.2.1 Proof of Claim 6

Proof. Pick any $A \in \hat{\mathfrak{F}}_2$ s.t. $\text{Pl}_2(\Theta_1|A) > 0$ (otherwise, $\text{Bl}_{\Theta_1}^{(\Theta_2)}(\cdot|A)$ does not exist). Clearly, $\text{Bl}_{\Theta_1}^{(\Theta_2)}(\emptyset|A) = 0$ and $\text{Bl}_{\Theta_1}^{(\Theta_2)}(\Theta_1|A) = 1$ by direct substitution. We also need to prove the following: for every positive integer n and every collection B_1, \dots, B_n s.t. $B_i \subseteq \Theta_1, i = \overline{1, n}$

$$\text{Bl}_{\Theta_1}^{(\Theta_2)}\left(\bigcup_{i=1}^n B_i \middle| A\right) \geq \sum_{\substack{I \subseteq \{1, \dots, n\}; \\ I \neq \emptyset}} (-1)^{|I|+1} \text{Bl}_{\Theta_1}^{(\Theta_2)}\left(\bigcap_{i \in I} B_i \middle| A\right). \quad (\text{A.20})$$

To show this, notice that, for any arbitrary set B , we have $\text{Bl}_{\Theta_2}(B|A) = \text{Bl}_{\Theta_2}(B \cap \Theta_2|A)$. Hence, without loss of generality, we can assume that $B_i \subseteq \Theta_1 \cap \Theta_2, i = \overline{1, n}$.

Since $\text{Bl}_{\Theta_2}(\cdot|A)$ is a valid belief function on Θ_2 , we must have

$$\text{Bl}_{\Theta_2}\left(\bigcup_{i=1}^n B_i \middle| A\right) \geq \sum_{\substack{I \subseteq \{1, \dots, n\}; \\ I \neq \emptyset}} (-1)^{|I|+1} \text{Bl}_{\Theta_2}\left(\bigcap_{i \in I} B_i \middle| A\right) \quad (\text{A.21})$$

Then the following two properties directly follow from the fact that $B_i \cap \Theta_{2 \setminus 1} = \emptyset, \forall i \in \{1, \dots, n\}$:

$$D_{\bigcup_{i=1}^n B_i} = \bigcup_{i=1}^n D_{B_i}; \quad D_{\bigcap_{i=1}^n B_i} = \bigcap_{i=1}^n D_{B_i}. \quad (\text{A.22})$$

Thus,

$$\begin{aligned} \text{Bl}_{\Theta_2}\left(D_{\bigcup_{i=1}^n B_i} \middle| A\right) &= \text{Bl}_{\Theta_2}\left(\bigcup_{i=1}^n D_{B_i} \middle| A\right) \\ &\geq \sum_{\substack{I \subseteq \{1, \dots, n\}; \\ I \neq \emptyset}} (-1)^{|I|+1} \text{Bl}_{\Theta_2}\left(\bigcap_{i \in I} D_{B_i} \middle| A\right) \\ &= \sum_{\substack{I \subseteq \{1, \dots, n\}; \\ I \neq \emptyset}} (-1)^{|I|+1} \text{Bl}_{\Theta_2}\left(D_{\bigcap_{i \in I} B_i} \middle| A\right) \end{aligned} \quad (\text{A.23})$$

Since $\sum_{\emptyset \neq I \subseteq \{1, \dots, n\}} (-1)^{|I|+1} = 1$, we also have

$$\text{Bl}_{\Theta_2}(\Theta_{2 \setminus 1} | A) = \sum_{\substack{I \subseteq \{1, \dots, n\}; \\ I \neq \emptyset}} (-1)^{|I|+1} \text{Bl}_{\Theta_2}(\Theta_{2 \setminus 1} | A). \quad (\text{A.24})$$

The claim follows by direct substitution of (A.21), (A.23) and (A.24) into (5.1) in the definition of $\text{Bl}_{\Theta_1}^{(\Theta_2)}(\cdot | A)$. \blacksquare

A.2.2 Proof of Claim 8

Proof. Clearly, $m_{\Theta_1}^{(\Theta_2)}(\emptyset | A) = \text{Bl}_{\Theta_1}^{(\Theta_2)}(\emptyset | A) = 0$. Now, let $\emptyset \neq B \subseteq \Theta_1$ be arbitrary. Expressing the belief terms in (5.1) in terms of masses and cross multiplying by $K_2(\Theta_1 | A)$, we get

$$\begin{aligned} & K_2(\Theta_1 | A) \cdot \sum_{C \subseteq B} m_{\Theta_1}^{(\Theta_2)}(C | A) \\ &= \sum_{C \subseteq B} m_{\Theta_2}(C | A) + \sum_{\substack{C \subseteq B; \\ D \subseteq \Theta_{2 \setminus 1}}} m_{\Theta_2}(C \cup D | A) - \sum_{D \subseteq \Theta_{2 \setminus 1}} m_{\Theta_2}(D | A) \\ &= \sum_{C \subseteq B} m_{\Theta_2}(C | A) + \sum_{\substack{\emptyset \neq C \subseteq B; \\ D \subseteq \Theta_{2 \setminus 1}}} m_{\Theta_2}(C \cup D | A) \\ &= \sum_{\emptyset \neq C \subseteq B} \left[m_{\Theta_2}(C | A) + \sum_{D \subseteq \Theta_{2 \setminus 1}} m_{\Theta_2}(C \cup D | A) \right]. \end{aligned} \quad (\text{A.25})$$

Since $m_{\Theta_2}(G | A) = 0$ for all $G \not\subseteq A$ (see [82] for details) and $m_{\Theta_1}^{(\Theta_2)}(\emptyset | A) = 0$, we get

$$\sum_{\emptyset \neq C \subseteq B} \left\{ m_{\Theta_1}^{(\Theta_2)}(C | A) - \frac{1}{K_2(\Theta_1 | A)} \left[m_{\Theta_2}(C | A) + \sum_{D \subseteq A \setminus \Theta_1} m_{\Theta_2}(C \cup D | A) \right] \right\} = 0. \quad (\text{A.26})$$

The claim then follows because $\emptyset \neq B \subseteq \Theta_1$ is arbitrary. \blacksquare

A.3 Chapter 6 Proofs

A.3.1 Proof of Claim 16

Proof. We need to show that $F_{\triangleleft}^i : \mathfrak{E}_{\Theta}^n \mapsto \mathfrak{E}_{\Theta}$ as given in Definition 34 has infinitely many fixed points and the core of a fixed point $\mathcal{E}_{\Theta} \in \text{fix}(F_{\triangleleft}^i)$ is s.t. for all $B \in \mathfrak{F}_{\Theta}$, $\nexists C \in \mathfrak{F}_{\Theta}$ s.t. $B \subset C$ or $B \supset C$. Let us proceed as follows.

Part 1: Here, we show that F_{\triangleleft}^i has infinitely many fixed points, viz., we need to show that there are infinite number of $\mathcal{E}_{\Theta} \in \mathfrak{E}_{\Theta}$ s.t. $\mathcal{E}_{\Theta} = F_{\triangleleft}^i(\mathcal{E}_{\Theta}, \dots, \mathcal{E}_{\Theta})$. Let us proceed as follows.

Construct a BoE $\mathcal{E}_{\Theta} = \{\Theta, \mathfrak{F}_{\Theta}, m_{\Theta}(\cdot)\}$ as follows. Pick \mathfrak{F}_{Θ} s.t. for all $B \in \mathfrak{F}_{\Theta}$, $\nexists C \in \mathfrak{F}_{\Theta}$ satisfying $B \subset C$ or $C \subset B$ for $B, C \subseteq \Theta$. Such a set exists, since $\mathfrak{F}_{\Theta} = \{B, \Theta \setminus B\}$, for any arbitrary $\emptyset \neq B \subset \Theta$ satisfies this condition. Now, pick any arbitrary mass assignment $m_{\Theta}(\cdot)$ s.t. $\sum_{A \in \mathfrak{F}_{\Theta}} m_{\Theta}(A) = 1$. Therefore, $\mathcal{E}_{\Theta} \in \mathfrak{E}_{\Theta}$. Now, we show that $\mathcal{E}_{\Theta} \in \text{fix}(F_{\triangleleft}^i)$.

By construction of \mathfrak{F}_{Θ} , according to CCT (Theorem 4) we have

$$m(B|C) = 0, \quad \text{for all } B, C \in \mathfrak{F}_{\Theta} \text{ s.t. } B \neq C. \quad (\text{A.27})$$

Now, let $\mathcal{E}'_{\Theta} = F_{\triangleleft}^i(\mathcal{E}_{\Theta}, \dots, \mathcal{E}_{\Theta})$, where $\mathcal{E}'_{\Theta} = \{\Theta, \mathfrak{F}'_{\Theta}, m'_{\Theta}(\cdot)\}$. So, the mass function for any $B \in \mathfrak{F}_{\Theta}$ gives

$$\begin{aligned} m'_{\Theta}(B) &= C_i m_{\Theta}(B) + \sum_{\substack{j=1; \\ j \neq i}}^n \sum_{A \in \mathfrak{F}_{\Theta}} C_j m_{\Theta}(A) m_{\Theta}(B|A) \\ &= C_i m_{\Theta}(B) + \sum_{\substack{j=1; \\ j \neq i}}^n C_j m_{\Theta}(B) m_{\Theta}(B|B), \quad \text{from (A.27)} \\ &= m_{\Theta}(B) \left(C_i + \sum_{\substack{j=1; \\ j \neq i}}^n C_j \right) \quad \text{from (4.2) (CCT)} \end{aligned}$$

$$= m_{\Theta}(B) \quad \text{from (6.14),}$$

where the last line follows from the fact that $\alpha_i + \sum \sum \beta_{ij} = 1$ in (6.14). Therefore, we have $m'_{\Theta}(B) = m_{\Theta}(B)$ for all $B \in \mathfrak{F}_{\Theta}$. Hence, we get $\mathfrak{F}'_{\Theta} = \mathfrak{F}_{\Theta}$ and also $m'_{\Theta}(\cdot) = m'_{\Theta}$. Therefore $\mathcal{E}'_{\Theta} = \mathcal{E}_{\Theta}$, and hence $\mathcal{E}_{\Theta} \in \text{fix}(F_{\triangleleft}^i)$. However, by construction, \mathfrak{F}_{Θ} is arbitrary and there are infinite number of $m_{\Theta}(\cdot)$, thus there are infinitely many $\mathcal{E}_{\Theta} \in \mathfrak{E}_{\Theta}$ satisfying above. Hence, F_{\triangleleft}^i has infinitely many fixed points.

Part 2: Let $\mathcal{E}_{\Theta} \in \text{fix}(F_{\triangleleft}^i)$. Now, we need to show that, for all $B \in \mathfrak{F}_{\Theta}$, $\nexists C \in \mathfrak{F}_{\Theta}$ s.t. $B \subset C$ or $B \supset C$. Let us proceed as follows.

Consider the mass function of F_{\triangleleft}^i for any arbitrary $B \in \mathfrak{F}_{\Theta}$. Since, $\mathcal{E}_{\Theta} \in \text{fix}(F_{\triangleleft}^i)$, we have

$$\begin{aligned} m_{\Theta}(B) &= C_i m_{\Theta}(B) + \sum_{\substack{j=1; \\ j \neq i}}^n \sum_{A \in \mathfrak{F}_{\Theta}} C_j m_{\Theta}(A) m_{\Theta}(B|A) \\ &= C_i m_{\Theta}(B) + \sum_{\substack{j=1; \\ j \neq i}}^n C_j \left(\sum_{A \in \mathfrak{F}_{\Theta}} m_{\Theta}(B) m_{\Theta}(B|A) \right) \\ &= C_i m_{\Theta}(B) + (1 - C_i) m_{\Theta}(B) \sum_{A \in \mathfrak{F}_{\Theta}} m_{\Theta}(B|A), \quad \text{from (6.14)} \\ \implies m_{\Theta}(B) &= m_{\Theta}(B) \sum_{A \in \mathfrak{F}_{\Theta}} m_{\Theta}(B|A), \quad \text{since } C_i \neq 1. \end{aligned}$$

Since, $B \in \mathfrak{F}_{\Theta}$ (viz., $m_{\Theta}(B) > 0$), we have $\sum_{A \in \mathfrak{F}_{\Theta}} m_{\Theta}(B|A) = 1$, for all $B \in \mathfrak{F}_{\Theta}$.

Then, use CCT to obtain the claim.

This completes the proof. ■

A.3.2 Proof of Claim 17

Proof. Consider n arbitrary BoEs $\mathcal{E}_{\Theta_i} \in \mathfrak{E}_{\Theta}$, $i = 1, \dots, n$ and an arbitrary fix point $\mathcal{E}_{\Theta} \in \text{fix}(F_{\triangleleft}^i)$. So, we need to prove $\|F_{\triangleleft}^i(\mathcal{E}_{\Theta_1}, \dots, \mathcal{E}_{\Theta_n}) - \mathcal{E}_{\Theta}\| < \max_j \|\mathcal{E}_{\Theta_j} - \mathcal{E}_{\Theta}\|$, or

otherwise $\mathcal{E}_{\Theta_j} = \mathcal{E}_{\Theta}$, $j = 1, \dots, n$. Let us proceed as follows.

Let $\mathfrak{B}_j \subseteq \mathfrak{F}_{\Theta_j}$ s.t. for all $B \in \mathfrak{B}_j$, $\nexists C \in \mathfrak{F}_{\Theta_j}$ s.t. $C \subset B$. Now, for any $B \in \mathfrak{B}_j$,

$$\begin{aligned} \sum_{A \in \mathfrak{F}_{\Theta_j}} m_{\Theta_j}(A) m_{\Theta_j}(B|A) &= \sum_{\substack{A \in \mathfrak{F}_{\Theta_j}; \\ A \supseteq B}} m_{\Theta_j}(A) m_{\Theta_j}(B|A) + \sum_{\substack{A \in \mathfrak{F}_{\Theta_j}; \\ A \subset B}} m_{\Theta_j}(A) \underbrace{m_{\Theta_j}(B|A)}_{=0 \text{ from CCT}} \\ &= m_{\Theta_j}(B) \underbrace{m_{\Theta_j}(B|B)}_{=1} + \sum_{\substack{A \in \mathfrak{F}_{\Theta_j}; \\ A \supset B}} m_{\Theta_j}(A) m_{\Theta_j}(B|A) \\ \implies \sum_{A \in \mathfrak{F}_{\Theta_j}} m_{\Theta_j}(A) m_{\Theta_j}(B|A) &\geq m_{\Theta_j}(B), \end{aligned} \quad (\text{A.28})$$

Since, $\sum_{B \subseteq \Theta} \sum_{A \in \mathfrak{F}_{\Theta_j}} m_{\Theta_j}(A) m_{\Theta_j}(B|A) = 1$, we have

$$\sum_{A \in \mathfrak{F}_{\Theta_j}} m_{\Theta_j}(A) m_{\Theta_j}(B|A) \leq m_{\Theta_j}(B), \quad \text{for all } B \notin \mathfrak{B}_j. \quad (\text{A.29})$$

Now, Consider \mathfrak{F}_{Θ} . As we have shown earlier, for all $B \in \mathfrak{F}_{\Theta}$, $\nexists C \in \mathfrak{F}_{\Theta}$ s.t. $B \subset C$ or $B \supset C$. Therefore, for all $B \in \mathfrak{F}_{\Theta}$, and for any $p \in \mathbb{R}^+$ s.t. $p \geq 1$ we have,

$$\sum_{B \in \mathfrak{F}_{\Theta}} \left| \sum_{A \in \mathfrak{F}_{\Theta_j}} m_{\Theta_j}(A) m_{\Theta_j}(B|A) - m_{\Theta}(B) \right|^p < \sum_{B \in \mathfrak{F}_{\Theta}} |m_{\Theta_j}(B) - m_{\Theta}(B)|^p, \quad (\text{A.30})$$

from (A.28), $\sum_{B \in \mathfrak{F}_{\Theta}} m_{\Theta}(B) = 1$ and the fact that $\mathcal{E}_{\Theta_j} \neq \mathcal{E}_{\Theta}$.

Now, for all $B \notin \mathfrak{F}_\Theta$, we have $\sum_{B \notin \mathfrak{F}_\Theta} m_\Theta(B) = 0$, hence

$$\begin{aligned}
& \sum_{B \notin \mathfrak{F}_\Theta} \left| \sum_{A \in \mathfrak{F}_{\Theta_j}} m_{\Theta_j}(A) m_{\Theta_j}(B|A) - m_\Theta(B) \right|^p \\
&= \sum_{B \notin \mathfrak{F}_\Theta} \left| \sum_{A \in \mathfrak{F}_{\Theta_j}} m_{\Theta_j}(A) m_{\Theta_j}(B|A) \right|^p, \quad \text{since } m_\Theta(B) = 0 \\
&\leq \sum_{B \notin \mathfrak{F}_\Theta} |m_{\Theta_j}(B)|^p, \quad \text{from (A.29)} \\
&= \sum_{B \notin \mathfrak{F}_\Theta} |m_{\Theta_j}(B) - m_\Theta(B)|^p, \quad \text{since } m_\Theta(B) = 0, \\
\Rightarrow \sum_{B \notin \mathfrak{F}_\Theta} \left| \sum_{A \in \mathfrak{F}_{\Theta_j}} m_{\Theta_j}(A) m_{\Theta_j}(B|A) - m_\Theta(B) \right|^p &\leq \sum_{B \notin \mathfrak{F}_\Theta} |m_{\Theta_j}(B) - m_\Theta(B)|^p. \quad (\text{A.31})
\end{aligned}$$

From (A.30) and (A.31), we get

$$\sum_{B \subseteq \Theta} \left| \sum_{A \in \mathfrak{F}_{\Theta_j}} m_{\Theta_j}(A) m_{\Theta_j}(B|A) - m_\Theta(B) \right|^p < \sum_{B \notin \mathfrak{F}_\Theta} |m_{\Theta_j}(B) - m_\Theta(B)|^p. \quad (\text{A.32})$$

We can now show the contracting property of F_\triangleleft^i as,

$$\begin{aligned}
& \|F_\triangleleft^i(\mathcal{E}_{\Theta_1}, \dots, \mathcal{E}_{\Theta_n}) - \mathcal{E}_\Theta\|^p \\
&= \sum_{B \subseteq \Theta} \left| C_i m_{\Theta_i}(B) + \sum_{\substack{j=1; \\ j \neq i}}^n \sum_{A \in \mathfrak{F}_{\Theta_j}} C_j m_{\Theta_j}(A) m_{\Theta_j}(B|A) - m_\Theta(B) \right|^p \\
&\leq \sum_{B \subseteq \Theta} C_i |m_{\Theta_i}(B) - m_\Theta(B)|^p \\
&\quad + \sum_{B \subseteq \Theta} (1 - C_i) \left| \sum_{A \in \mathfrak{F}_{\Theta_j}} \sum_{\substack{j=1; \\ j \neq i}}^n \frac{C_j}{1 - C_i} m_{\Theta_j}(A) m_{\Theta_j}(B|A) - m_\Theta(B) \right|^p \\
&= \sum_{B \subseteq \Theta} C_i |m_{\Theta_i}(B) - m_\Theta(B)|^p \\
&\quad + \sum_{B \subseteq \Theta} (1 - C_i) \left| \sum_{\substack{j=1; \\ j \neq i}}^n \frac{C_j}{1 - C_i} \sum_{A \in \mathfrak{F}_{\Theta_j}} m_{\Theta_j}(A) m_{\Theta_j}(B|A) - m_\Theta(B) \right|^p
\end{aligned}$$

$$\begin{aligned}
&\leq \sum_{B \subseteq \Theta} C_i |m_{\Theta_i}(B) - m_{\Theta}(B)|^p \\
&\quad + \max_k \sum_{B \subseteq \Theta} (1 - C_i) \left| \sum_{\substack{j=1; \\ j \neq i}}^n \frac{C_j}{1 - C_i} \sum_{A \in \mathfrak{F}_{\Theta_k}} m_{\Theta_k}(A) m_{\Theta_k}(B|A) - m_{\Theta}(B) \right|^p
\end{aligned}$$

since, $\sum_{j \neq i} C_j = 1 - C_i$, we get

$$\begin{aligned}
&\|F_{\triangleleft}^i(\mathcal{E}_{\Theta_1}, \dots, \mathcal{E}_{\Theta_n}) - \mathcal{E}_{\Theta}\|^p \\
&\leq \sum_{B \subseteq \Theta} C_i |m_{\Theta_i}(B) - m_{\Theta}(B)|^p \\
&\quad + \max_k \sum_{B \subseteq \Theta} (1 - C_i) \left| \sum_{A \in \mathfrak{F}_{\Theta_k}} m_{\Theta_k}(A) m_{\Theta_k}(B|A) - m_{\Theta}(B) \right|^p
\end{aligned}$$

using (A.32), we get

$$\begin{aligned}
&\|F_{\triangleleft}^i(\mathcal{E}_{\Theta_1}, \dots, \mathcal{E}_{\Theta_n}) - \mathcal{E}_{\Theta}\|^p \\
&< \sum_{B \subseteq \Theta} C_i |m_{\Theta_i}(B) - m_{\Theta}(B)|^p + \max_k \sum_{B \subseteq \Theta} (1 - C_i) |m_{\Theta_k}(B) - m_{\Theta}(B)|^p \\
&= \max_k \sum_{B \subseteq \Theta} \left\{ C_i |m_{\Theta_i}(B) - m_{\Theta}(B)|^p + (1 - C_i) |m_{\Theta_k}(B) - m_{\Theta}(B)|^p \right\} \\
&= \max_k \sum_{B \subseteq \Theta} |m_{\Theta_i}(B) - m_{\Theta}(B)|^p
\end{aligned}$$

Therefore,

$$\begin{aligned}
\|F_{\triangleleft}^i(\mathcal{E}_{\Theta_1}, \dots, \mathcal{E}_{\Theta_n}) - \mathcal{E}_{\Theta}\| &< \max_k \left[\sum_{B \subseteq \Theta} |m_{\Theta_i}(B) - m_{\Theta}(B)|^p \right]^{\frac{1}{p}} \\
&= \max_k \|m_{\Theta_i}(B) - m_{\Theta}(B)\|.
\end{aligned}$$

This completes the proof. ■

A.3.3 Proof of Claim 19

Proof. Here, we need prove that the pool $\mathcal{F}_{\triangleleft} = \{F_{\triangleleft}^i \mid i \in \mathbb{I}\}$ is paracontractive on \mathfrak{E}_{Θ} and contains infinitely many common fixed-points in \mathfrak{E}_{Θ} . This is very easily established as follows.

Construct a BoE \mathcal{E}_{Θ} with an arbitrary \mathfrak{F}_{Θ} s.t. for all $B \in \mathfrak{F}_{\Theta}$, $\nexists C \in \mathfrak{F}_{\Theta}$ s.t. $B \subset C$ or $C \subset B$. Now, let $m_{\Theta}(\cdot)$ be arbitrary mass assignment s.t. $\sum_{B \in \mathfrak{F}_{\Theta}} m_{\Theta}(B) = 1$. Now, as we have already shown, \mathcal{E}_{Θ} is a fixed point of F_{\triangleleft}^i for all $i \in \mathbb{I}$, there for it's a common fixed-point of \mathfrak{F}_{Θ} . ■

Bibliography

- [1] J. Llinas, C. Bowman, G. L. Rogova, A. Steinberg, E. Waltz, and F. White, “Revisiting the JDL Data Fusion Model II,” in *Proc. International Conference on Information Fusion (ICIF'04)*, Stockholm, Sweden, Jun./Jul. 2004, pp. 1218–1230.
- [2] L. A. Zadeh, “Review of shafer’s mathematical theory of evidence,” *Artificial Intelligence Magazine*, vol. 5, pp. 81–83, 1984.
- [3] M. C. Florea, A. Josselme, E. Bossé, and D. Grenier, “Robust combination rules for evidence theory,” *Information Fusion*, vol. 10, no. 2, pp. 183 – 197, 2009.
- [4] F. E. White, “Data Fusion Lexicon, Joint Directors of Laboratories, Technical Panel for C³,” Naval Ocean Systems Center, San Diego, Tech. Rep., 1987.
- [5] (2001) Reshaping the military for asymmetric warfare. Center for Defense Information. [Online]. Available: <http://www.cdi.org/terrorism/asymmetric.cfm>
- [6] M. Pravia, R. Prasanth, P. Arambel, C. Sidner, and C.-Y. Chong, “Generation of a fundamental data set for hard/soft information fusion,” in *Information Fusion, 2008 11th International Conference on*, jul 2008, pp. 1 –8.
- [7] E. Waltz and J. Llinas, *Multisensor Data Fusion*. Norwood, MA: Artech House, 1990.
- [8] V. V. S. Sarma and S. Raju, “Multisensor data fusion and decision support for airborne target identification,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 21, no. 5, Sep./Oct. 1991.

- [9] D. L. Hall, *Mathematical Techniques in Multisensor Data Fusion*. Norwood, MA: Artech House, 1992.
- [10] J. W. Carl, "Contrasting approaches to combine evidence," in *Handbook of Multisensor Data Fusion*, ser. The Electrical Engineering and Applied Signal Processing Series, D. L. Hall and J. Llinas, Eds. Boca Raton, FL: CRC Press, 2001, pp. 7.1–7.32.
- [11] D. L. Hall and J. Llinas, Eds., *Handbook of Multisensor Data Fusion*. Boca Raton, FL: CRC Press, 2001.
- [12] G. Seetharaman, H. V. Le, S. S. Iyengar, and R. Logananthraj, "SmartSAM: A multisensor network based framework for video surveillance and monitoring," in *Proc. International Symposium on Mathematical Theory of Networks and Systems (MTNS'04)*, Leuven, Belgium, Jul. 2004, Session WA7.
- [13] D. Smith and S. Singh, "Approaches to multisensor data fusion in target tracking: A survey," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 18, no. 12, pp. 1696–1710, Dec 2006.
- [14] K. Premaratne, M. N. Murthi, J. Zhang, M. Scheutz, and P. H. Bauer, "A Dempster-Shafer theoretic conditional approach to evidence updating for fusion of hard and soft data," in *Proc. International Conference on Information Fusion (ICIF'09)*, Seattle, WA, Jul. 2009, pp. 2122–2129.
- [15] M. Thomas, "Model based fusion and services automation support," in *Information Fusion, 2009. FUSION '09. 12th International Conference on*, Jul. 2009, pp. 1600–1603.
- [16] D. Hall, M. M. J., Llinas, and T. Mullen, "A framework for dynamic hard/soft fusion," in *Information Fusion, 2008 11th International Conference on*, July 2008, pp. 1–8.
- [17] K. Sambhoos, J. Llinas, and E. Little, "Graphical methods for real-time fusion and estimation with soft message data," in *Proc. International Conference on Information Fusion (ICIF'08)*, Cologne, Germany, Jun./Jul. 2008, pp. 1–8.
- [18] C. Laudy and B. Goujon, "Soft data analysis within a decision support system," in *Proc. International Conference on Information Fusion (ICIF'09)*, Seattle, WA, Jul. 2009, pp. 1889–1896.
- [19] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory sensing," in *In: Workshop on World-Sensor-Web (WSW06): Mobile Device Centric Sensor Networks and Applications*, 2006, pp. 117–134.

- [20] R. Lucky, "Reflections: A billion amateurs," *Spectrum, IEEE*, vol. 44, no. 11, p. 96, Nov 2007.
- [21] F. Crestani and G. Pasi, "Soft information retrieval: applications of fuzzy set theory and neural networks," in *Neuro-fuzzy Techniques for Intelligent Information Systems*. Physica Verlag (Springer Verlag), 1999, pp. 287–315.
- [22] G. Ferrin, L. Snidaro, S. Canazza, and G. Foresti, "Soft data issues in fusion of video surveillance," in *Information Fusion, 2008 11th International Conference on*, Jul 2008, pp. 1–8.
- [23] M. A. Pravia, O. Babko-Malaya, M. K. Schneider, J. V. White, C.-Y. Chong, and A. S. Willsky, "Lessons learned in the creation of a data set for hard/soft information fusion," in *Proc. International Conference on Information Fusion (ICIF'09)*, Seattle, WA, Jul. 2009, pp. 2114–2121.
- [24] B. Khaleghi, A. Khamis, and F. Karray, "Random finite set theoretic based soft/hard data fusion with application for target tracking," in *Multisensor Fusion and Integration for Intelligent Systems (MFI), 2010 IEEE Conference on*, Sep. 2010, pp. 50–55. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5604458
- [25] R. Yager, "Hard and soft information fusion using measures," in *Intelligent Systems and Knowledge Engineering (ISKE), 2010 International Conference on*, Nov 2010, pp. 13–16.
- [26] M. Webster, *Merriam Websters Collegiate Dictionary*. Merriam Webster, July 2003.
- [27] R. Kennes and P. Smets, "Fast algorithms for Dempster-Shafer theory," in *Proc. 3rd International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'90)*, ser. Lecture Notes in Computer Science, vol. 521. Paris, France: Springer-Verlag, Jul 1991, pp. 14–23.
- [28] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press, 1976.
- [29] P. Smets, "Probability, possibility, belief: Which and where?" in *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, ser. Quantified Representation of Uncertainty and Imprecision, D. Gabbay and P. Smets, Eds. Dordrecht, Netherlands: Kluwer, 1998, vol. 1, pp. 1–24.
- [30] H. Borotschnig, L. Paletta, M. Prantl, and A. Pinz, "A comparison of probabilistic, possibilistic and evidence theoretic fusion schemes for active object recognition," *Computing*, vol. 62, no. 4, pp. 293–319, 1999.

- [31] F. Delmotte, "Comparison of the performances of decision aimed algorithms with Bayesian and beliefs basis," *International Journal of Intelligent Systems*, vol. 16, pp. 963–981, 2001.
- [32] R. R. Yager, M. Fedrizzi, and J. Kacprzyk, Eds., *Advances in the Dempster-Shafer Theory of Evidence*. New York, NY: John Wiley and Sons, 1994.
- [33] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Norwood, MA: Artech House, 1999.
- [34] J. R. Boston, "A signal detection system based on Dempster-Shafer theory and comparison to fuzzy detection," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 30, no. 1, pp. 45–51, Feb. 2000.
- [35] H. Altincay and M. Demirekler, "Speaker identification by combining multiple classifiers using Dempster-Shafer theory of evidence," *Speech Communication*, vol. 41, no. 4, pp. 531–547, Nov. 2003.
- [36] T. M. Chen and V. Venkataramanan, "Dempster-Shafer theory for intrusion detection in ad hoc networks," *IEEE Internet Computing*, vol. 9, no. 6, pp. 35–41, Nov./Dec. 2005.
- [37] D. A. Dewasurendra, P. H. Bauer, and K. Premaratne, "Evidence filtering," *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5796–5805, 2007.
- [38] P. L. Bogler, "Shafer-Dempster reasoning with applications to multisensor target identification systems," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 17, no. 6, pp. 968–977, 1987.
- [39] H. Leung and J. Wu, "Bayesian and Dempster-Shafer target identification for radar surveillance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 2, pp. 432–447, Apr. 2000.
- [40] F. Delmotte and P. Smets, "Target identification based on the transferable belief model interpretation of Dempster-Shafer model," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 34, no. 4, pp. 457–471, Jul. 2004.
- [41] S. L. Hegarat-Masclé, I. Bloch, and D. Vidal-Madjar, "Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 1018–1031, Jul. 1997.
- [42] S. S. Anand, D. A. Bell, and J. G. Hughes, "EDM: A general framework for data mining based on evidence theory," *Data and Knowledge Engineering*, vol. 18, pp. 189–223, 1996.

- [43] D. Cai, M. F. McTear, and S. I. McClean, "Knowledge discovery in distributed databases using evidence theory," *International Journal of Intelligent Systems*, vol. 15, no. 8, pp. 745–761, Jun. 2000.
- [44] Z. Elouedi, K. Mellouli, and P. Smets, "Belief decision trees: theoretical foundations," *International Journal of Approximate Reasoning*, vol. 28, no. 2/3, pp. 91–124, Nov. 2001.
- [45] H. Altincay, "A Dempster-Shafer theoretic framework for boosting based ensemble design," *Pattern Analysis and Applications Journal*, vol. 8, no. 3, pp. 287–302, Dec. 2005.
- [46] K. K. R. G. K. Hewawasam, K. Premaratne, and M.-L. Shyu, "Rule mining and classification in a situation assessment application: a belief theoretic approach for handling data imperfections," *IEEE Transaction on Systems, Man and Cybernetics, Part B: Cybernetics*, vol. 37, no. 6, pp. 1446–1459, Dec. 2007.
- [47] K. K. R. G. K. Hewawasam and K. Premaratne, "Dependency based reasoning in a Dempster-Shafer theoretic framework," in *Proc. International Conference on Information Fusion (ICIF'07)*, Quebec, Canada, Jul. 2007, pp. 1–8.
- [48] T. L. Wickramarathne, K. Premaratne, M. Kubat, and D. T. Jayaweera, "CoFiDS: A belief theoretic approach for automated collaborative filtering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, pp. 175–189, May 2011.
- [49] ICIF, *International Conference on Information Fusion (ICIF'08)*. Cologne, Germany: International Society of Information Fusion, Jul. 2008.
- [50] —, *International Conference on Information Fusion (ICIF'09)*. Seattle, WA: International Society of Information Fusion, Jul. 2009.
- [51] —, *International Conference on Information Fusion (ICIF'08)*. Edinburgh, UK: International Society of Information Fusion, Jul. 2010.
- [52] T. L. Wickramarathne, K. Premaratne, M. N. Murthi, and M. Scheutz, "A Dempster-Shafer theoretic evidence updating strategy for non-identical frames of discernment," in *Proc. Workshop on the Theory of Belief Functions (WTBF'10)*, Brest, France, Apr 2010.
- [53] T. L. Wickramarathne, K. Premaratne, and M. N. Murthi, "Focal elements generated by the Dempster-Shafer theoretic conditionals: A complete characterization," in *Proc. International Conference on Information Fusion (ICIF'10)*, Scotland, UK, Jul. 2010.

- [54] D. Lewis, “Probabilities of conditionals and conditional probabilities,” *The Philosophical Review*, vol. LXXXV, no. 3, pp. 297–315, Jul. 1976.
- [55] A. Benavoli, L. Chisci, A. Farina, and B. Ristic, “Modelling uncertain implication rules in evidence theory,” in *Proc. International Conference on Information Fusion (ICIF’08)*, Cologne, Germany, Jun./Jul. 2008, pp. 1–7.
- [56] L. Chrisman, “Incremental conditioning of lower and upper probabilities,” *International Journal of Approximate Reasoning*, vol. 13, no. 1, pp. 1–25, Jul. 1995.
- [57] S. C. A. Thomopoulos, “Sensor Integration and Data Fusion,” *Journal of Robotics Systems*, vol. 7, pp. 337–372, 1990.
- [58] P. P. Bonissone and R. M. Tong, “Editorial: Reasoning with uncertainty in expert systems,” *International Journal of Man-Machine Studies*, vol. 22, no. 3, pp. 241–250, 1985.
- [59] P. Smets, “Varieties of ignorance and the need for well-founded theories,” *Information Sciences*, vol. 57-58, pp. 135–144, September 1991.
- [60] S. Das, *High-Level Data Fusion*. Norwood, MA, USA: Artech House, Inc., 2008.
- [61] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann, 1988.
- [62] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.
- [63] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer, 1997.
- [64] L. A. Zadeh, “Fuzzy sets as a basis for a theory of possibility,” *Fuzzy Sets and Systems*, vol. 100, Supplement 1, no. 0, pp. 9 – 34, 1999.
- [65] D. Dubois and H. Prade, “An introduction to possibilistic and fuzzy logics,” in *Readings in Uncertain Reasoning*, G. Shafer and J. Pearl, Eds. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990, pp. 742–761.
- [66] —, “Possibility theory as a basis for qualitative decision theory,” in *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2*, ser. IJCAI’95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995, pp. 1924–1930.

- [67] P. Smets, "Practical uses of belief functions," in *Proc. Conference on Uncertainty in Artificial Intelligence (UAI'99)*, K. B. Laskey and H. Prade, Eds. San Francisco, CA: Morgan Kaufmann, 1999, pp. 612–621.
- [68] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Annals of Mathematical Statistics*, vol. 38, no. 2, pp. 325–339, Apr. 1967.
- [69] R. Fagin and J. Y. Halpern, "A new approach to updating beliefs," in *Proc. Conference on Uncertainty in Artificial Intelligence (UAI'91)*, P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, Eds. New York, NY: Elsevier Science, 1991, pp. 347–374.
- [70] A. P. Dempster, "A generalization of Bayesian inference," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 30, no. 2, pp. 205–247, 1968.
- [71] R. R. Yager, "On the dempster-shafer framework and new combination rules," *Information Sciences*, vol. 41, no. 2, pp. 93 – 137, 1987.
- [72] F. Smarandache and J. Dezert, "A simple proportional conflict redistribution rule," *JAIF Journal*, pp. 1–36, 2005.
- [73] F. Janez and A. Appriou, "Theory of evidence and non-exhaustive frames of discernment: Plausibilities correction methods," *International Journal of Approximate Reasoning*, vol. 18, no. 1/2, pp. 1–19, Jan./Feb. 1998.
- [74] K. Premaratne, D. A. Dewasurendra, and P. H. Bauer, "Evidence combination in an environment with heterogeneous sources," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 37, no. 3, pp. 298–309, 2007.
- [75] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.
- [76] A.-L. Jousselme, D. Grenier, and E. Bosse, "A new distance between two bodies of evidence," *Information Fusion*, vol. 2, no. 2, pp. 91–101, Jun. 2001.
- [77] V. Kimala and K. Yamada, "Another conditional rule in evidence theory," in *Proc. Conference on Humans and Computers (CHC)*, Nagaoka, Japan, Nov 2008.
- [78] R. Fagin and J. Y. Halpern, "Uncertainty, belief and probability," *Computational Intelligence*, vol. 7, pp. 160–173, 1991.
- [79] P. Diaconis and S. L. Zabell, "Some alternatives to Bayes' rule," in *Proc. Second University of California, Irvine, Conference on Political Economy*, 1986, pp. 25–38.

- [80] P. Diaconis, “Review of [A Mathematical Theory of Evidence. (Glenn Shafer)],” *Journal of the American Statistical Association*, vol. 73, no. 363, pp. 677–678, 1978.
- [81] B. Ristic and P. Smets, “Target identification using belief functions and implication rules,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 3, pp. 1097–1103, July 2005.
- [82] E. C. Kulasekere, K. Premaratne, D. A. Dewasurendra, M.-L. Shyu, and P. H. Bauer, “Conditioning and updating evidence,” *International Journal of Approximate Reasoning*, vol. 36, no. 1, pp. 75–108, Apr. 2004.
- [83] M. Pott, “On the convergence of asynchronous iteration methods for nonlinear paracontractions and consistent linear systems,” *Linear Algebra and its Applications*, vol. 283, no. 13, pp. 1 – 33, 1998.
- [84] R. Olfati-Saber and J. S. Shamma, “Consensus filters for sensor networks and distributed sensor fusion,” in *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC ’05. 44th IEEE Conference on*, dec 2005, pp. 6698 – 6703.
- [85] L. Xiao, S. Boyd, and S. Lall, “A scheme for robust distributed sensor fusion based on average consensus,” in *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, april 2005, pp. 63 – 70.
- [86] L. Schenato and G. Gamba, “A distributed consensus protocol for clock synchronization in wireless sensor network,” in *Decision and Control, 2007 46th IEEE Conference on*, dec 2007, pp. 2289 –2294.
- [87] M. Cao, A. S. Morse, and B. D. O. Anderson, “Reaching a consensus in a dynamically changing environment:a graphical approach,” *SIAM Journal of Control and Optimization*, vol. 47, no. 2, pp. 575–600, 2008.
- [88] L. Fang and P. J. Antsaklis, “Asynchronous consensus protocols using nonlinear paracontractions theory,” *IEEE Transactions on Automatic Control*, vol. 53, no. 10, pp. 2351 –2355, Nov 2008.
- [89] S. Kar and J. M. F. Moura, “Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise,” *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 355 –369, jan 2009.
- [90] Z. Wu, H. Fang, and Y. She, “Weighted average prediction for improving consensus performance of second-order delayed multi-agent systems,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. PP, no. 99, pp. 1–9, 2012.

- [91] M. Pott, *On the convergence of asynchronous iteration methods for nonlinear paracontractions and consistent linear systems*. Sonderforschungsbereich 343, 1995.
- [92] K. Lehrer and C. Wagner, *Rational Consensus in Science and Society*, ser. Philosophical studies series in philosophy. D. Reidel Pub. Co., 1981.
- [93] A. Martin, A.-L. Jousselme, and C. Osswald, "Conflict measure for the discounting operation on belief functions," in *Information Fusion, 2008 11th International Conference on*, July 2008.
- [94] P. Walley, *Statistical Reasoning with Imprecise Probabilities*. London, UK: Chapman and Hall, 1991.
- [95] J. Y. Jaffray, "Bayesian updating and belief functions," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, no. 5, pp. 1144–1152, Oct. 1992.
- [96] N. Wilson, "Algorithms for dempster-shafer theory," in *Algorithms for Uncertainty and Defeasible Reasoning*. Kluwer Academic Publishers, 2000, pp. 421–475.
- [97] F. Voorbraak, "A computationally efficient approximation of Dempster-Shafer theory," *International Journal of Man-Machine Studies*, vol. 30, no. 5, pp. 525–536, 1989.
- [98] B. Tessem, "Approximations for efficient computation in the theory of evidence," *Artificial Intelligence*, vol. 61, no. 2, pp. 315 – 329, 1993.
- [99] J. D. Lowrance, T. D. Garvey, and T. M. Strat, "A framework for evidential reasoning systems," in *Proc. National Conference of the American Association for Artificial Intelligence (AAAI'86)*, 1986, pp. 896–903.
- [100] M. Bauer, "Approximations for decision making in the Dempster-Shafer theory of evidence—an empirical study," in *Proc. Conference on Uncertainty in Artificial Intelligence (UAI)*, P. P. Bonissone, Ed. Elsevier Science, Aug./Oct. 1997, vol. 17, no. 2/3, pp. 217–237.
- [101] N. Wilson, "A Monte-Carlo algorithm for Dempster-Shafer belief," in *Uncertainty in Artificial Intelligence*, 1991, pp. 414–417.
- [102] M. Clarke and N. Wilson, "Efficient algorithms for belief functions based on the relationship between belief and probability," in *Symbolic and Quantitative Approaches to Uncertainty*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 1991, vol. 548, pp. 48–52.

- [103] N. Wilson, “The combination of belief: when and how fast?” *International Journal of Approximate Reasoning*, vol. 6, pp. 377–388, May 1992.
- [104] V. Savchuk and H. Martz, “Bayes reliability estimation using multiple sources of prior information: binomial sampling,” *Reliability, IEEE Transactions on*, vol. 43, no. 1, pp. 138–144, Mar. 1994.

